Data Economy & Innovation Cycles: An International Comparison*

December 2, 2021

Man Cho

The KDI School of Public Policy & Mgt.

(* This presentation is based on the findings documented by Cho·Moon·Rhee·Choi (2021))

Bio of Man CHO

Man CHO is a full professor at KDI School of Public Policy and Management in Korea, and is currently serving as Chairman of International Academy of Financial Consumers, an international academic association with over 100 members from more than 20 countries. His teaching and research areas include FinTech, credit risk management, real estate finance, and and urban and regional economics. Before joining KDI School in May 2007, he worked for Fannie Mae, one of the major MBS (Mortgage Backed Security) issuers in the U.S., in which he involved with various R&D projects such as mortgage and MBS pricing, collateral (property) assessment, and mortgage default and prepayment modeling and also served several managerial positions. Prior to Fannie Mae, he worked for the World Bank as a long-term consultant (1991~92), and taught at the Johns Hopkins University as an adjunct professor (2004~07). He holds a Ph.D degree in Applied Economics and Managerial Science from the Wharton School of the University of Pennsylvania in 1991.

Abstract

The main objective of this research is to conceptualize the term “data economy (DE),” and, further, to examine the issue of how to operationalize that in three particular service sectors in Korea – finance, real estate, and medical service through an international comparison. To that end, we surveyed the recent studies of relevancy, and also performed the meta-analyses by utilizing two sets of the DE-related academic studies in the three sectors that are extracted from two international databases: 263,974 articles on the health and medical services from PubMed established by the National Institute of Health in the U.S.; and, 26,859 articles on the finance and real estate from the Web of Science. Our results show that: DE is defined as an ecosystem that can enhance social welfare in a national or regional level through accumulating·sharing·analyzing•utilizing digital data; Its three core elements include a DPA(Data•Platform•AI) based infrastructure along with institutional and market driven stimulators; Its expected socio-economic effects as argued in the literature are categorized into three types - (1) the platform effect, (2) the predictive power effect, and (3) the new analytics effect. And the results of the meta-analyses indicate that: The number of academic studies on DE related to the three sectors has been in an explosive growth path since around 2015; The co-authorship and the inter-country collaboration have also been increasing steadily; As to the outcomes achieved by the Korean researchers, the country fairs reasonably well in terms of quantity (ranked as 8th place for both databases) but tends to lag in terms of quality (the average number of citations, and the newness of topics). Using the above results, several policy implications are discussed.

Outline

- Introduction
- Conceptualizing Data Economy (DE)
- DE: Its Expected Social Effects
- The Meta Analyses Performed
- Policy Implications
- Concluding Remarks

Introduction

Digital Transformation (DT): The governmental responses

- Strategizing the economic social development in response to the on-going DT trend
  - USA: The OPEN Government Data Act of 2019 (initiating OpenData initiative & Data.gov
  - China: The plan for developing BigData industry (2017), & developing the Data Exchange
  - EU: The EU data privercy directive (EU GDPR, 2018); Initiating GAIA-X project, to activate 'data economy' (2018)
  - Japan: Reform for Society 5.0 (2017) ~ to establish the institutional base for data economy and manpower development plan
  - UK: The UK Digital Strategy (2017), and the national data strategy (2020)

Digital Transformation (DT): Its impact on the global industry

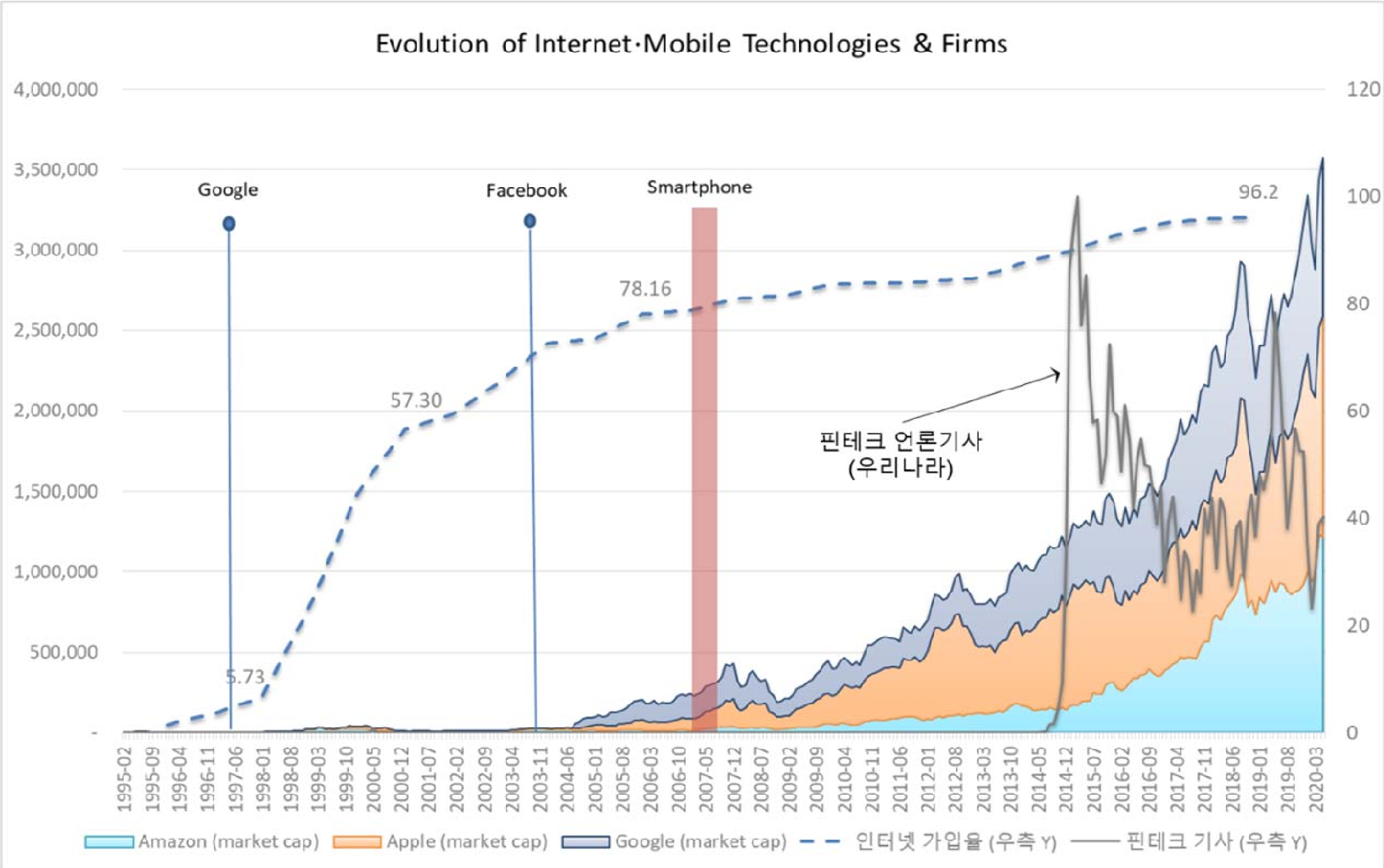
- The dynamic re-structuring of the global industrial structure
  - Change of the global top 10 companies (2007~2017)
  - Rise of the US & Chinese global companies, but decline of the European big corporations
    - Changes in the top 100 global firms (last 20 years): US ~ 50 → 61; China ~ 0 → 15; Europe ~ 41 → 15 (no change for the Korean & Japanese companies)
    - Trend of the global GDP: US ~ 24% (in 2020); China ~ 18%; Europe ~ 34% → 25% (btw 2000-2020)
    - The creative destruction processes being initiated by the US & Chines companies backed by Data·Platform•AI (DPA) ➔ To be termed as the DE-enabled industry

The global top 10 companies (Change in 2007~2017)

28-Jun-07		08-Sep-17			28-Jun-07			08-Sep-17
		1	Apple		1	Exxon		
		2	Alphabet (Google)		2	GE		
3		3	Microsoft		4	PetroChina		
		4	Facebook		5	Royal Dutch Shell		
		5	Amazon		6	Citigroup		
		6	Alibaba		7	AT&T		
		7	Tencent		8	Gazprom		
					9	BP		
		13	Samsung Electronics		10	Toyota		10
29								21
								24
70								27
81								30
367								38
Below 500								41
								69
								239

출처: 김희수 (2018)

[Data] Evolution of technologies, internet firms, and the use of ‘FinTech’ terminology



Research Questions

- How can we conceptualize 'Data Economy' (i.e., its definition, elements, market and institutional enablers, and so on)?
- What are its expected socio-economic outcomes, in the cases of three particular service sectors in Korea (i.e., finance, real estate, and medical service)?
- What trends do we observe from the academic research communities (related to the above sectors) as to DE?
- What policy implications can we draw from the analyses on the above question, in particular, for Korea?

Conceptualizing Data Economy (DE)

Data Economy: What is it?

Data Economy: Definition

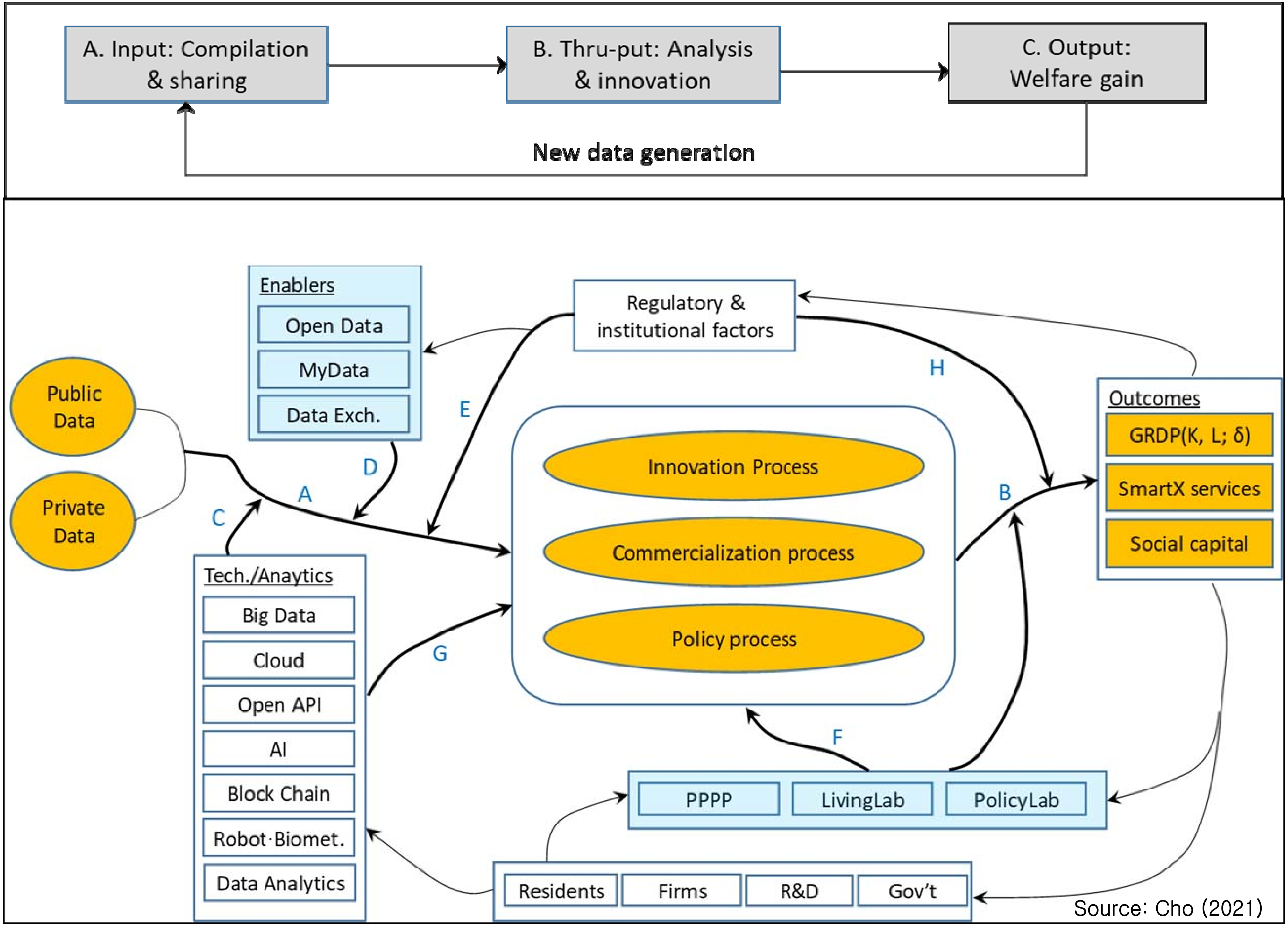
- An innovation ecosystem that generates socio-economic values by efficiently distributing or trading digital data

"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" (European Commission (2018))

Data Economy: Key elements

- Input: To accumulate & share digital data (BigData•Cloud•API•DLT)
- Thruput: To make data-driven innovations possible (AI & data analytics, ML & DL)
- Output: To increase social welfare with the data-driven innovations
- Plus various market and institutional "enablers"

A Basic Structure of Data Economy



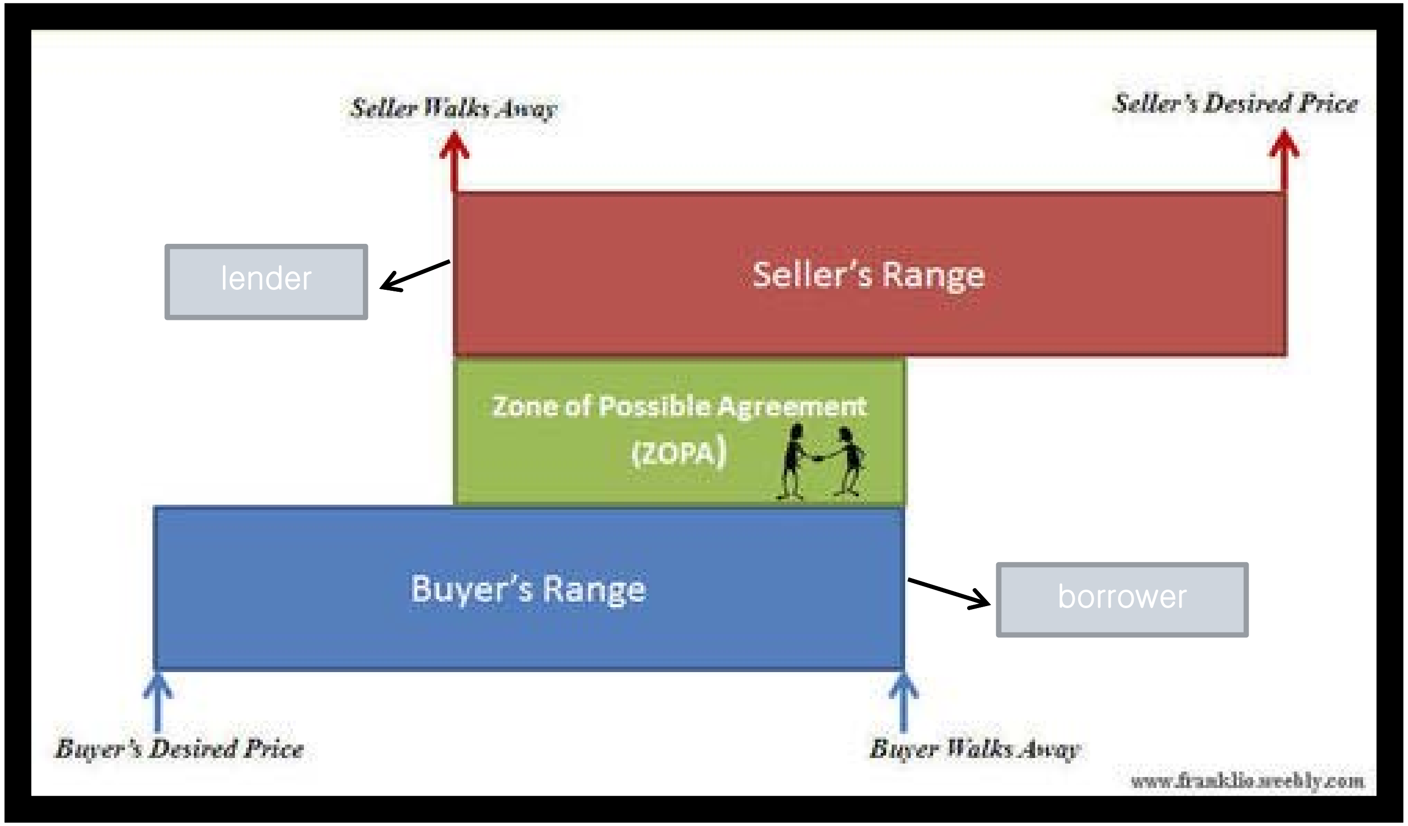
The DE-based Innovation Sectors

- The DE-based innovation sectors
  - Manufacturing: SmartFactory, SmartFarm, Driverless Car
  - Service, private: Digital health (e.g., online medical service), FinTech & InsurTech, PropTech
  - Service, public: SmartCity, Open Government

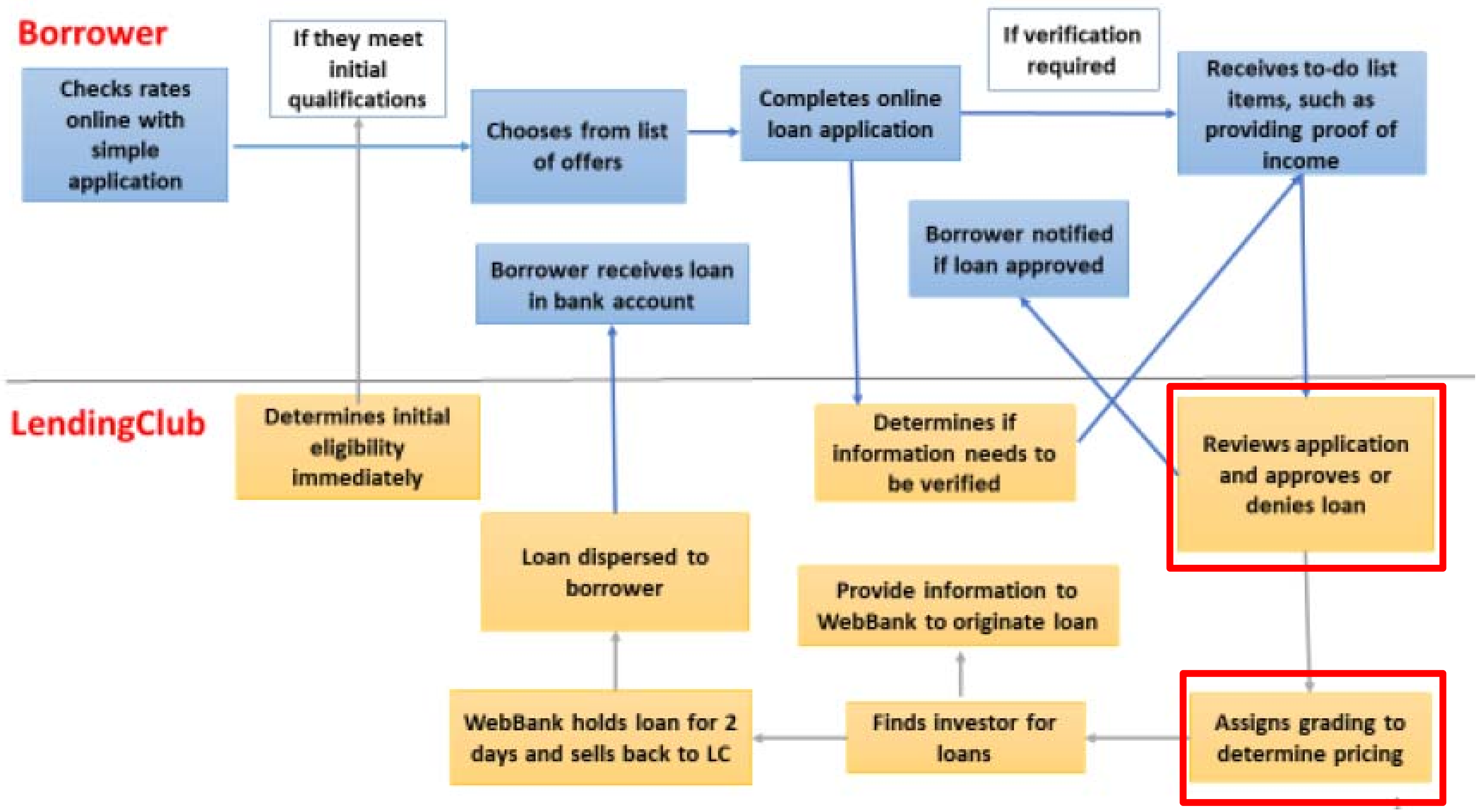
- Several cases of the DE- (or DPA-) based innovation
  - ZOPA (Zone of Possible Agreement)
  - Online credit evaluation by P2P lenders
  - SmartHome and prevention of Alzheimer disease
  - InsurTech and financial inclusion

Zone Of Possible Agreement;

(a negotiating term identifying the bounds within which agreement can be reached between two parties)

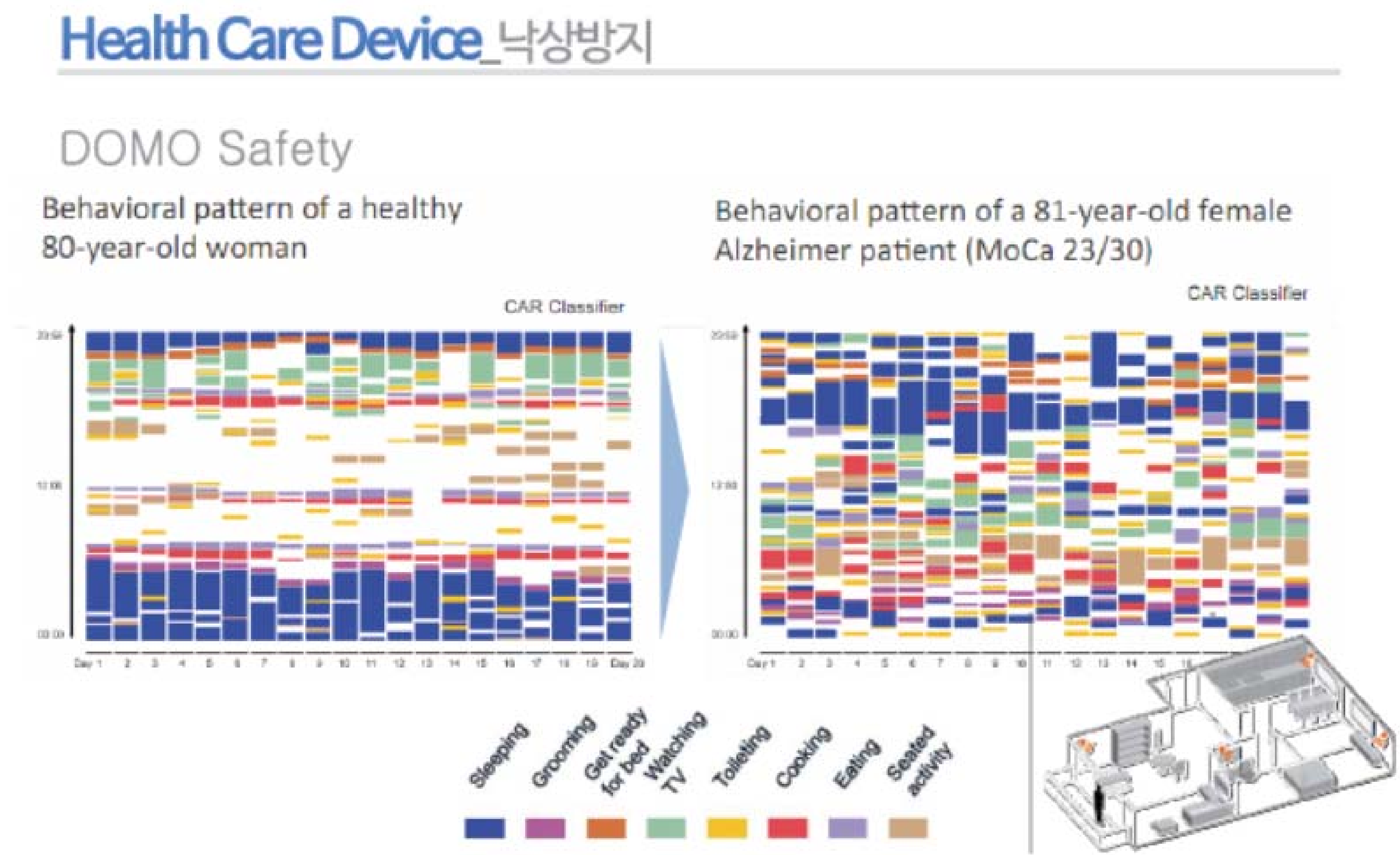


[참고] FinTech loan production process (the case of Lending Club)



Source: Jagtiani and Lemieux (2019)

InsurTech & Auto/Health Insurance

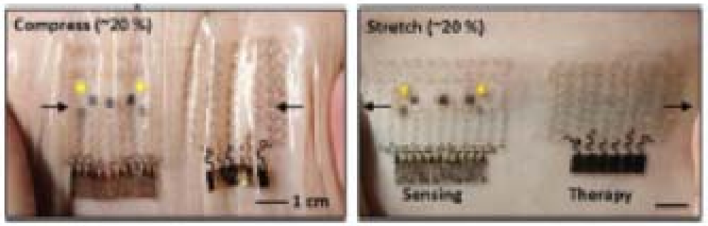


Wearables & InsurTech

Health Care Device_고혈압, 당뇨 관리

■ 웨어러블 혈당측정/투약

- 그래핀 전자피부



■ 웨어러블 연속혈당측정계

- Abbott 'Freestyle Libre'

- Dexcom G5

DE: Its Expected Social Effects

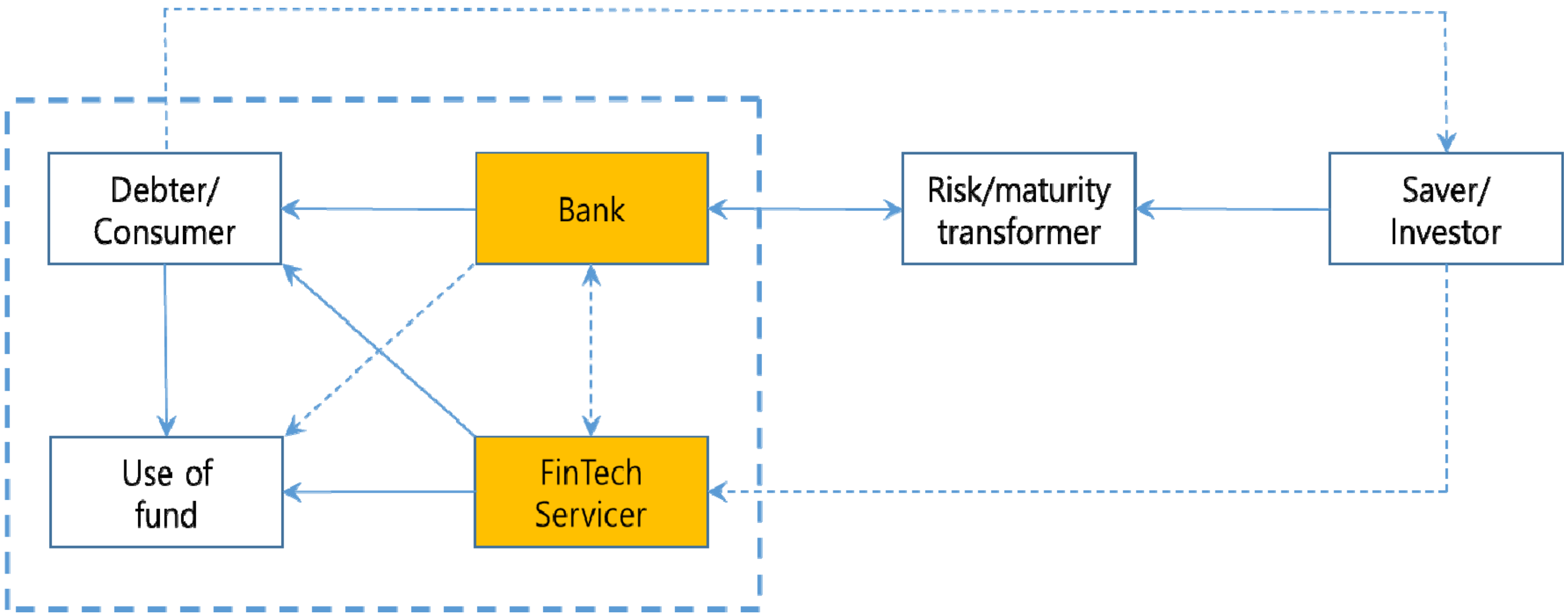
Socio-economic Effects of DE (from the Global Research Community)

- The rising trend of academic research on DE during the last several years ➔ Three main socio-economic effects of DE
  1. Platform effect
  2. Prediction power effect
  3. New (data) analytics effect
- Key arguments being summarized in the subsequent slides

1. Platform effect of DE

- WWW (1989) & smart phone (2007) ➔ Exponential growth of internet-mobile platforms, resulting in:
  - Dramatic reduction in transaction cost (TC)
  - Rise in digital data via online transactions (on consumer behavior)
- Implications to the three sectors
  - Finance: Rise of “disintermediation” (Fuster et al. (2020), Cho (2020), Philippon (2015) and (2016))
  - Real estate: The sector with high search cost; Rise of PropTech → TC↓ (Baum (2017))
  - Medical service: Rise of remote/inhouse medical service via platform → TC↓ (Oh (2020), Kang (2020))

[Data] Traditional & FinTech-based Financial Intermediation



출처: Cho (2020)

2. Predictive power effect of DE

- Accumulate·shate·utilize digital data ➔ Predictive power for optimal technology or business model (Farboodi/ Veldkamp (2021))
  - ① Digital data↑ → "MSE" in predicting optimal technology↓ → Quality-adjusted production↑ (individual firms)
  - ② Market-wide DE effect: Generally decreasing returns to scale (DRS); But IRS for start-ups & for the state of low data accumulation
  - ③ MSE↓ vs. Technology frontier(game changer)↑ ~ Another DE effect
- Related micro & macro DE effects
  - DE effect in predicting business cycles (Ordonez (2013), Fajgelbaum et al. (2017)), productivity growth rates (Agrawal et al. (2018))
  - Reduction in user cost of capital (Begenau et al. (2018)), and in the role of collateral in the lending sector (Gambacorta et al. (2020))

[Data] Key arguments by Farboodi/Veldkamp (2021)

$$(1)\ \ y_{i,t} = A_{i,t} \cdot k_{i,t}^{\alpha}$$

$$(2)\ \ A_{i,t} = \bar{A} - (a_{i,t} - \theta_t - \varepsilon_{a,i,t})^2$$

- y: Quality-adjusted output (firm i time t), per k units of capital
- A: Technology-driven quality (of goods/services produced); A bar ~ technology frontier (time and firm invariant); a ~ the firm's production technology choice
- θ, ε: Persistent and transitory components of the technology ($\theta_t + \varepsilon_{a,i,t}$); θ ~ AR(1) process with a random innovation
- The role of data: **Helping firms to choose better production techniques**, i.e., reducing the prediction error $\varepsilon_{a,i,t}$ such that **y↑ for given levels of factors (k, l)** or **new and improved business models or processes**
- Three key features of data: 1) Data is a by-product of economic activity; 2) data is information used for prediction, and 3) uncertainty reduction enhances firm profitability
- More fundamental data-driven innovation (i.e., **game changer**): Shifting the technology frontier (A bar)!

[Data] Key arguments by Farboodi/Veldkamp (2021) (cont'd)

- Diminishing returns for a data-accumulation economy: Over time, the aggregate stock of knowledge and aggregate amount of output would have a time path that resembles the concave path in Figure 2 (Without idea creation, data accumulation alone would generate slower and slower growth ➔ the Solow's growth path)
- Increasing returns to scale (the data feedback loop): A firm with more data produces higher-quality goods, which induces them to invest (on data) more, produce more, and sell more ➔ Causing aggregate knowledge accumulation to accelerate.
  - Diminishing returns always dominate when data is abundant; But when firms are young, or data is scarce, increasing returns can be strong enough to create an increasing rate of growth.
- Implications to our three sectors
  - Medical service: Data-driven innovations in developing new & better medicines or more efficient (and more preventive) curing methods
  - Financial service: Data-driven innovations in developing more efficient (& accurate) risk-assessment and decision makings (underwriting – or segmentation, pricing, and hedging) and more welfare-enhancing financial products or service processes
  - Real estate industry: Similar data-driven innovations as in finance (e.g., real estate FinTech); Other real estate specific innovations (SmartHome, RE-share economy, smart construction, and so on)

[Data] Key arguments by Farboodi/Veldkamp (2021) (cont'd)

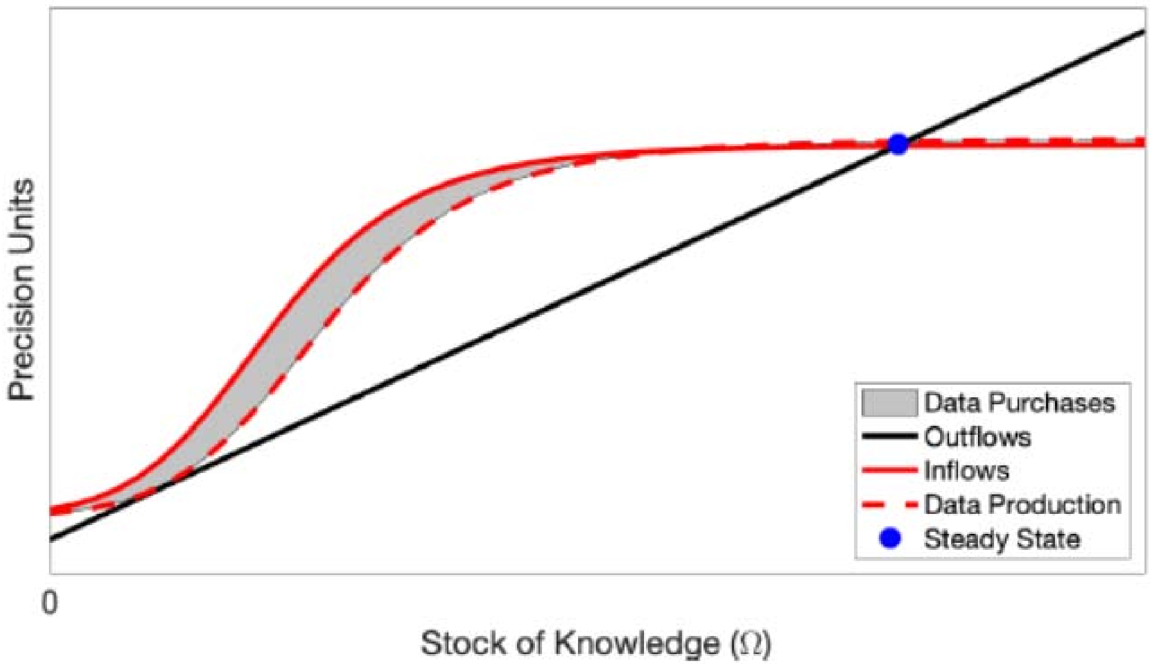


Figure 3: New firms grow slowly: inflows and outflows of data of a single firm.
Line labeled inflows plots the individual firm i version of the quantity in equation (10), that makes an optimal capital decision $k^*_{i,t}$ and data decision $\delta^*_{i,t}$, with different levels of initial data stock. This firm is in an economy where all other firms are in steady state. Line labeled outflows plots the individual firm i version of the quantity in (11). Data production is $z_i k^*_{i,t}{}^{\alpha} \sigma_\epsilon^{-2}$, which is inflows without the data purchases $\delta_{i,t}$.

2. Predictive power effect of DE (cont'd)

- Implications to the three sectors
  - Medical: Accumulation-utilization of medical data ➔ New medical procedures & medicine via R&D↑
  - Finance: Efficiency in risk measurement & mgt. (e.g., FinTech)↑
    - ① In measuring systematic & idiosyncratic risks; In credit evaluation
    - ② Use of alternative data (e.g., "digital footprint") (Berg et al. (2018) 등)
    - ➔ Information asymmetry↓ & financial inclusion↑
  - Real Estate: RE FinTech vs. non-finance PropTech
    - Non-finance RE: Shared economy; SmartConstruction (MSE↓)
    - RE market transparency↑: "Optimal" price indices (on the methodologies, serial and cross-sectional aggregations)

3. New (data) analytics effect of DE

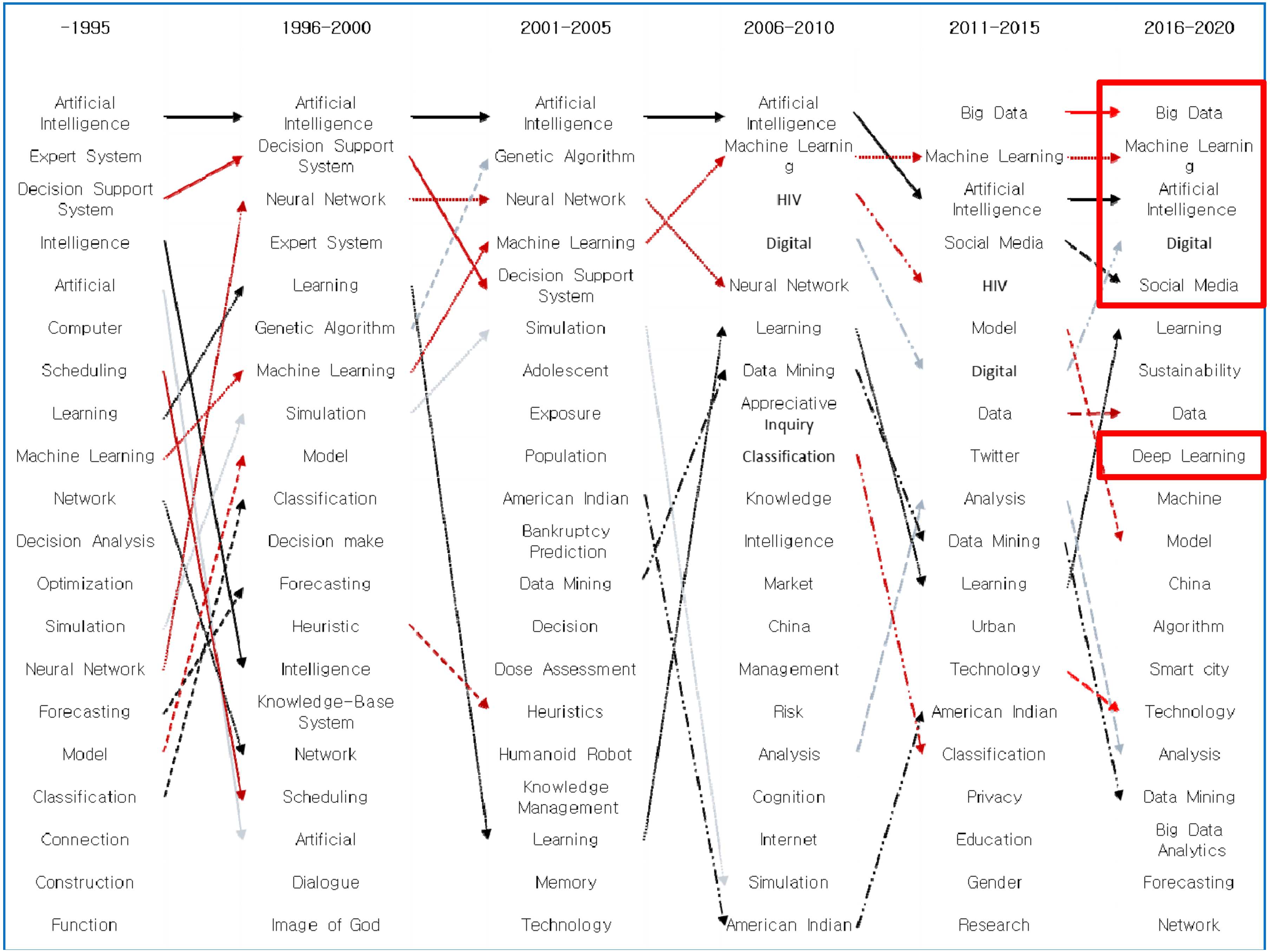
- Explosive growth of the AI-based data analytics
  - The AI transition (algorithms → Machine Learning (ML) → Deep Learning(DL)): Being accelerated during the last five years
  - Enabling the use of non-conventional data (e.g., measuring economic activity via satellite pictures, classifying industries via tax filings)
  - A new family of analytics (e.g., regression trees, LASSO, random forests, ensemble) & **the enabling software** (e.g., R, Python)
  - Key implications (Mullainathan/Spiess (2017)): (1) Focusing on predicting **dependent variable (y hat) rather than beta**; (2) Machine- learning on **numerous interactions & splines** among x variables; (3) Relying on **out-of-sample model performance tests**

$$y = \beta \cdot x + \varepsilon \rightarrow \hat{y} = \hat{\beta} \cdot x + \hat{\varepsilon}$$

The Trend of the PubMed MeSH Headings (91~20): Medical Research

Rank	1991-1995	1996-2000	2001-2005	2006-2010	2011-2015	2016-2020
1	Neural Networks, Computer	Neural Networks, Computer	Neural Networks, Computer	Algorithms	Algorithms	Machine Learning
2	Expert Systems	Artificial Intelligence	Algorithms	Artificial Intelligence	Neural Networks, Computer	Neural Networks, Computer
3	Artificial Intelligence	Image Processing, Computer-Assisted	Artificial Intelligence	Neural Networks, Computer	Artificial Intelligence	Algorithms
4	Image Processing, Computer-Assisted	Algorithms	Robotics	Robotics	Robotics	Deep Learning
5	Diagnosis, Computer-Assisted	Expert Systems	Signal Processing, Computer-Assisted	Signal Processing, Computer-Assisted	Support Vector Machine	Artificial Intelligence
6	Signal Processing, Computer-Assisted	Computer Simulation	Models, Biological	Models, Biological	Software	Robotics
7	Computer Simulation	Signal Processing, Computer-Assisted	User-Computer Interface	Subtraction Technique	Models, Theoretical	Software
8	Radiographic Image Enhancement	Robotics	Models, Neurological	User-Computer Interface	Natural Language Processing	Support Vector Machine
9	Software	Diagnosis, Computer-Assisted	Software	Natural Language Processing	Models, Biological	Models, Theoretical
10	Algorithms	Radiographic Image Enhancement	Image Processing, Computer-Assisted	Software	Signal Processing, Computer-Assisted	Transcriptome
11	Angiography, Digital Subtraction	Models, Neurological	Fuzzy Logic	Models, Neurological	Models, Neurological	Natural Language Processing
12	Radiology Information Systems	Software	Computer Simulation	Fuzzy Logic	Computer Simulation	Models, Biological
13	Models, Biological	Angiography, Digital Subtraction	Models, Statistical	Models, Theoretical	User-Computer Interface	Signal Processing, Computer-Assisted
14	Robotics	Fuzzy Logic	Subtraction Technique	Computer Simulation	Fuzzy Logic	Gene Expression Profiling
15	Magnetic Resonance Imaging	Tomography, X-Ray Computed	Natural Language Processing	Models, Statistical	Machine Learning	Magnetic Resonance Imaging
16	Decision Making, Computer-Assisted	Radiology Information Systems	Pattern Recognition, Automated	Nonlinear Dynamics	Models, Statistical	Computer Simulation
17	Models, Neurological	User-Computer Interface	Models, Theoretical	Decision Support Techniques	Internet	Image Processing, Computer-Assisted
18	Microcomputers	Models, Biological	Internet	Artifacts	Gene Expression Profiling	Electronic Health Records
19	Tomography, X-Ray Computed	Natural Language Processing	Angiography, Digital Subtraction	Image Processing, Computer-Assisted	Electronic Health Records	Internet
20	Decision Support Techniques	Computer Communication Networks	Expert Systems	Laparoscopy	Stroke Rehabilitation	Models, Neurological
21	Medical Records Systems, Computerized	Magnetic Resonance Imaging	Imaging, Three-Dimensional	Internet	Image Processing, Computer-Assisted	Phylogeny

The trend of WOS keywords: Finance & RE



The Meta Analyses Performed

The Meta Analyses: The DB compiled

- PubMed article Bibliography Data (medical science)
  - Keywords - Artificial Intelligence, Machine Learning, Digitalization, Digital Health (linked with "or"); Publication Date - 1862-2020
  - Data Cleaning (written in English, published in 1991 or later, and so on): 263,974 articles included (1991-2020)

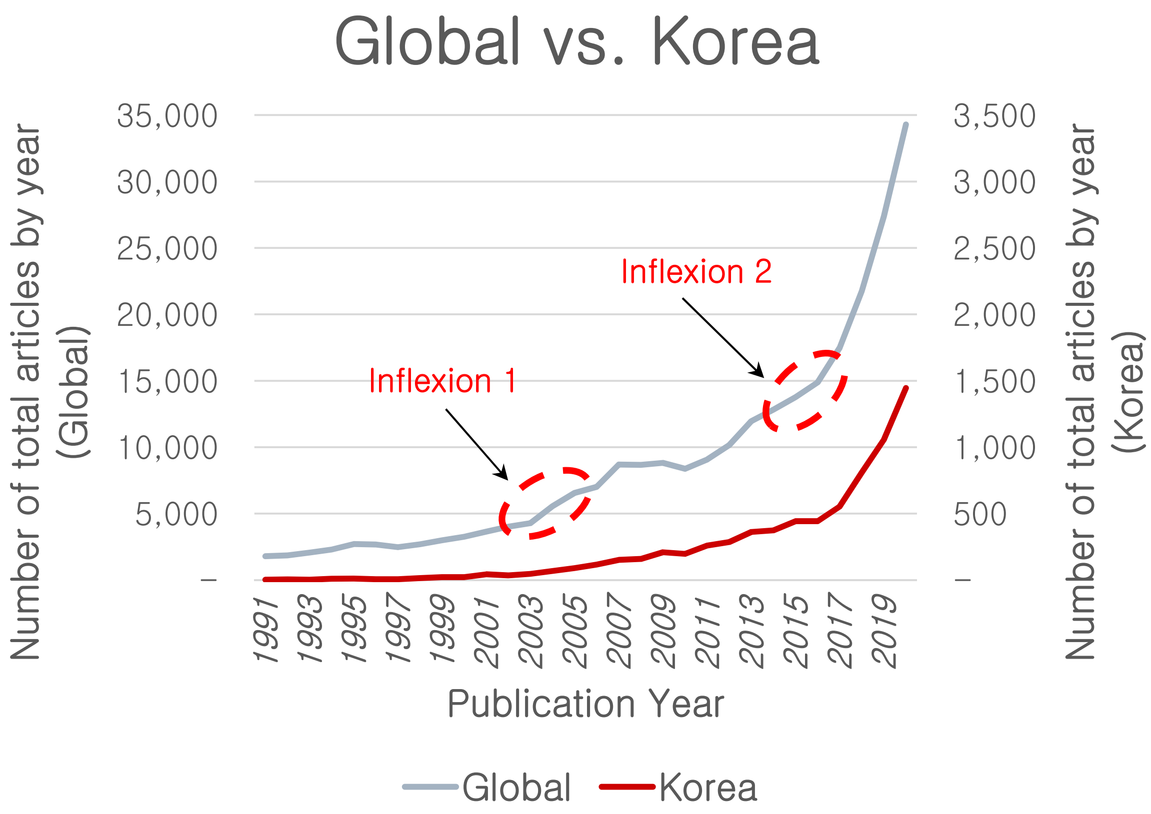
- Web of Science, WOS (finance, real estate, and related social science disciplines)
  - Keywords – FinTech (& its variations), PropTech, Data Economy, Big Data, AI, ML, DL, Digitalization ((linked with "or")
  - Scope of the journals – Selected among SSCI and some inter-disciplinary journals
  - Result – 26,859 articles included (1978~2022)

Observation (1): Explosive growth of the related research since around 2015

- PubMed articles (medical science): Frequency of the articles having explosive growth since around 2015
  - Two inflexion points globally – one around 2003 & another around 2015; Korea – only one around 2015
  - The internet (www) based innovations would have been more prevalent in the US and UK (further investigation needed)
- WOS articles (finance and RE):
  - Only one pronounced inflexion point around 2015, and an explosive growth since then
  - The social science research would be lagging that of natural science; The applied research (finance & RE) would be lagging the industry (i.e., research follows FinTech & PropTech)

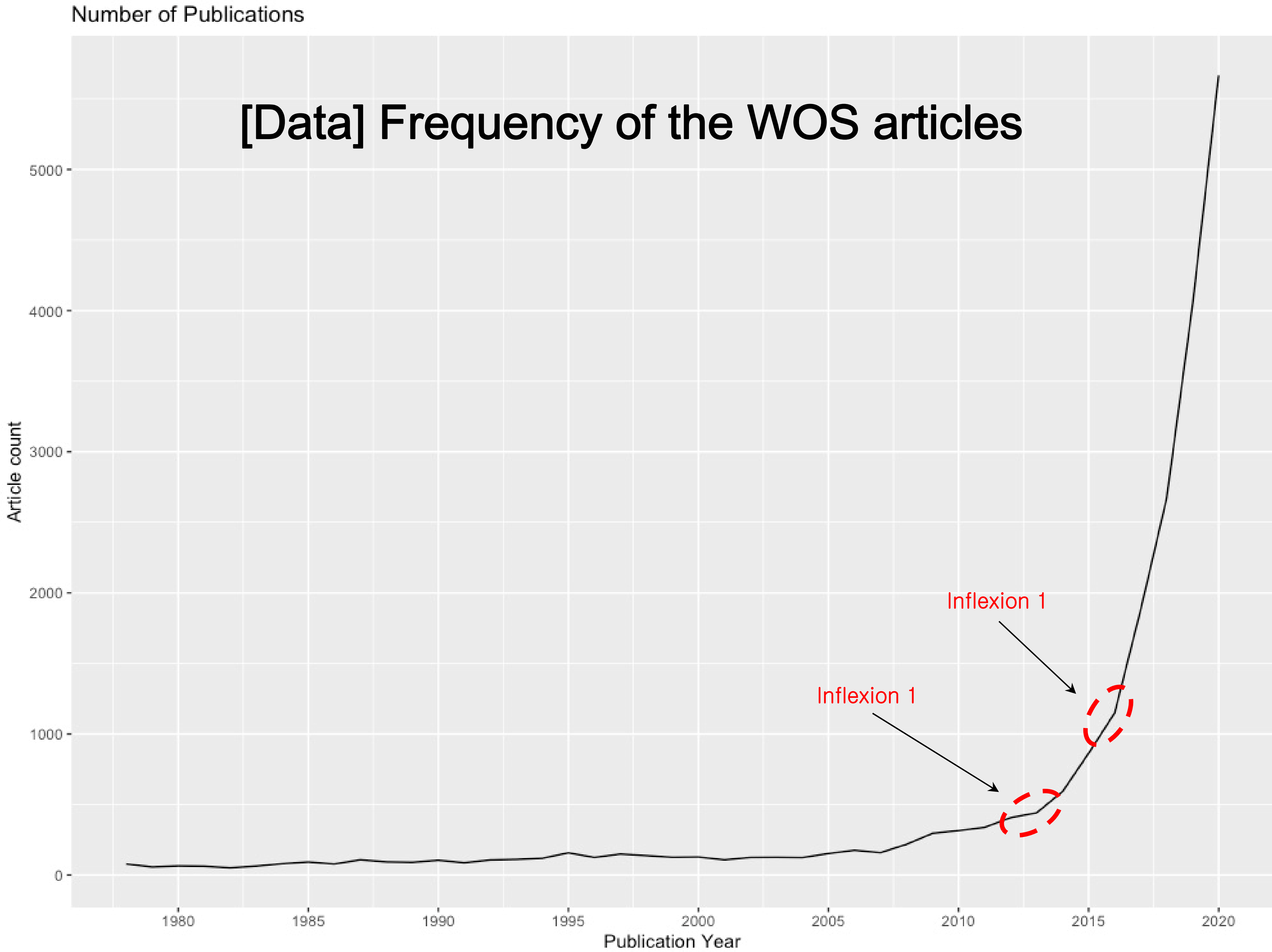
[Data] Frequency of the PubMed articles

Digitalization in Biotechnology and Medicine



Publication Year	Count (Article)	
	Global	Korea
1991	1,784	2
1992	1,847	4
1993	2,069	2
1994	2,296	9
1995	2,699	10
1996	2,672	6
1997	2,478	6
1998	2,692	14
1999	2,998	21
2000	3,260	21
2001	3,642	43
2002	4,011	34
2003	4,282	46
2004	5,559	68
2005	6,553	90
2006	7,016	116
2007	8,693	151
2008	8,662	158
2009	8,816	209
2010	8,370	197
2011	9,073	260
2012	10,162	286
2013	11,960	362
2014	12,830	374
2015	13,774	442
2016	14,871	442
2017	17,485	553
2018	21,753	809
2019	27,372	1,056
2020	34,295	1,445

[Data] Frequency of the WOS articles

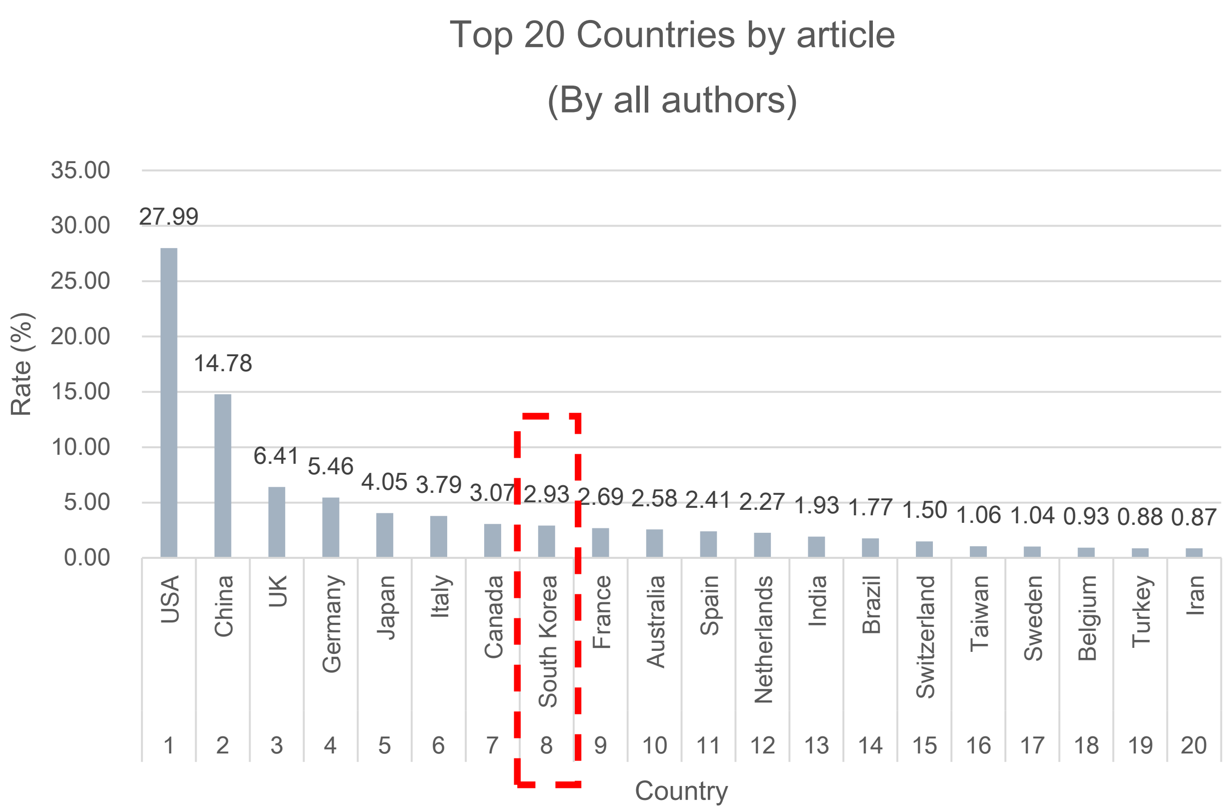


Observation (2): Dominance in research by three countries – US, China, and UK

- The variations observed across the top countries (PubMed)
  - US and China (UK to some degree) dominating the DE-related medical research (w/ close to a 50% of all articles), followed by Germany, Japan, & Italy
  - Korea being ranked as 8th place; A similar growth rate with other fast-growing countries in the research activity (a low initial level but a very fast growth, a la the log transformation)
- The variations observed across the top countries (WOS)
  - The same three countries (US, China, & UK) dominating the DE-related SS research (w/ close to a 50% of all articles), followed by Spain, Italy, and Germany
  - Korea being ranked as 8th place; A reasonably high-growth country

[Data] Distribution of the PubMed articles

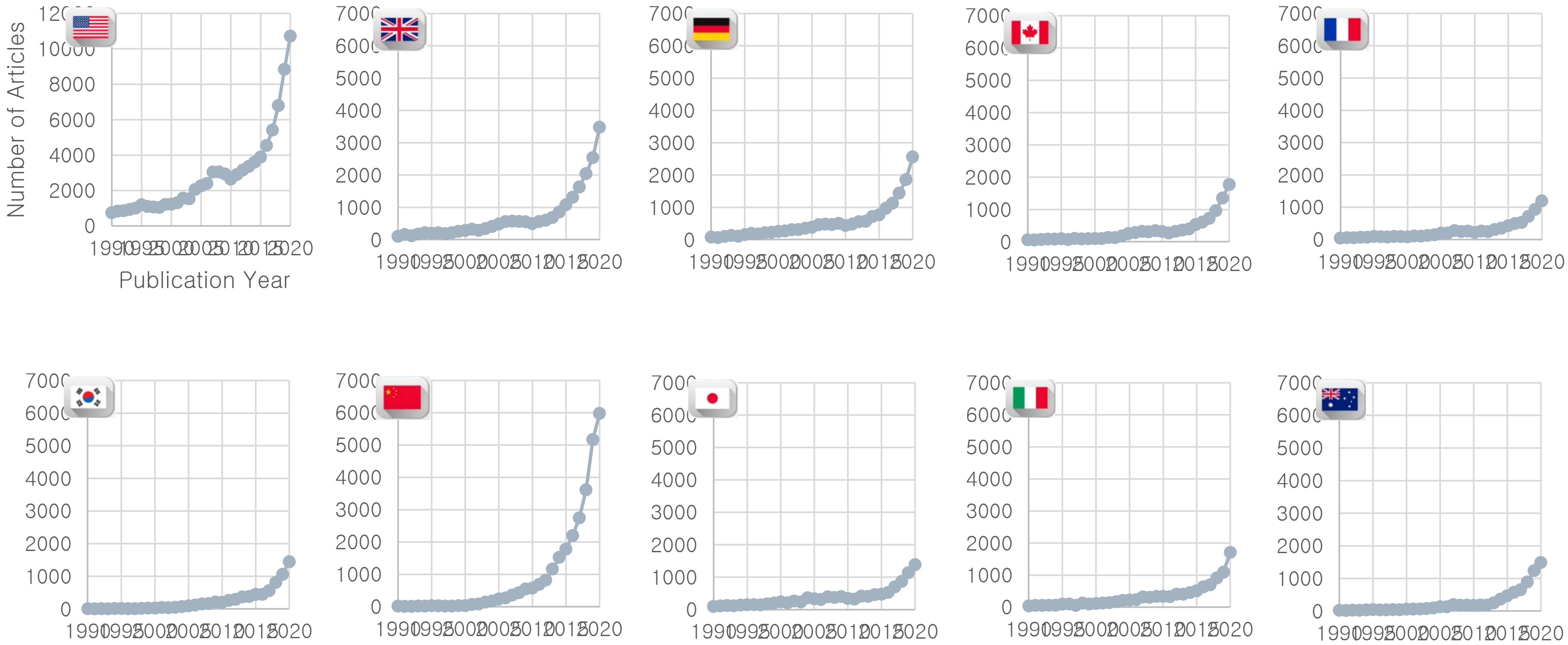
Top 20 Countries



	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

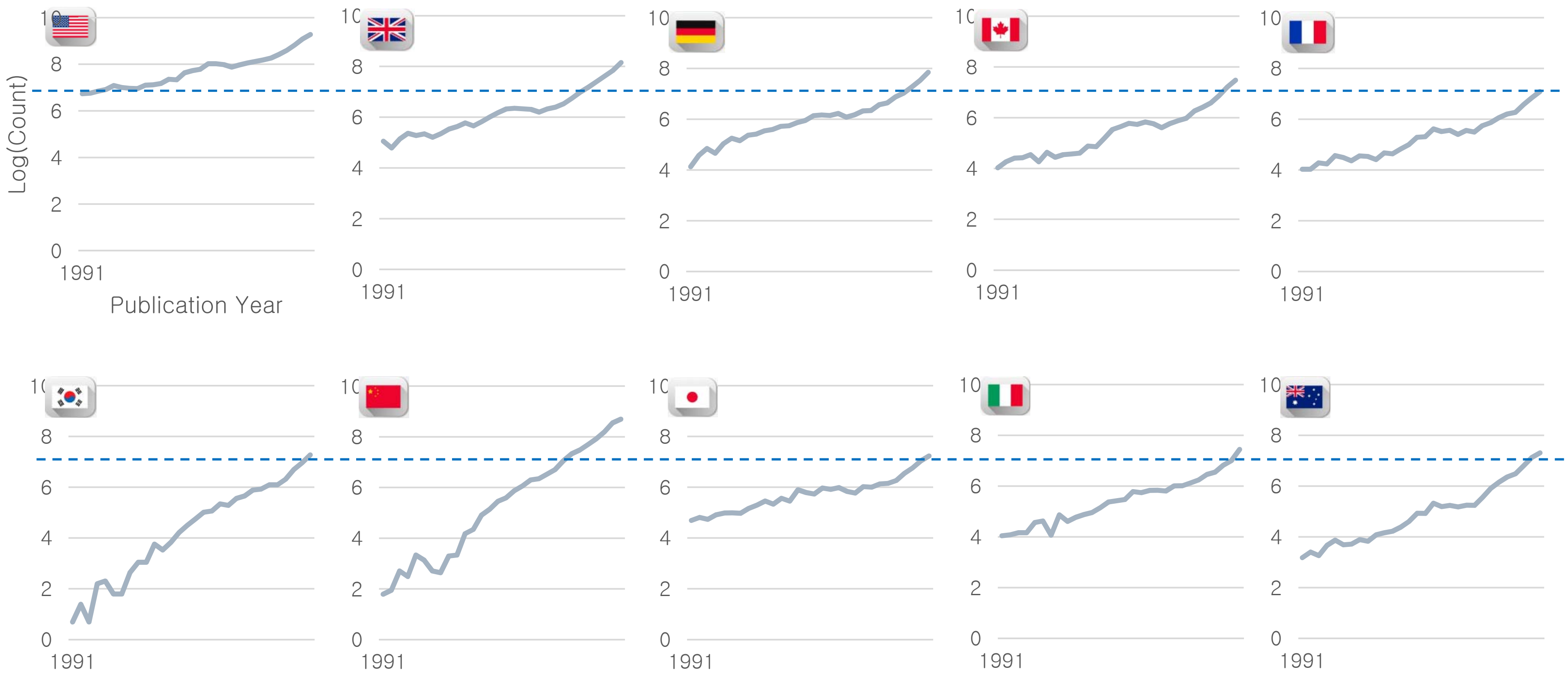
[Data] Growth Patterns, by Country

Growth in Numbers: 10 Countries



[Data] Growth Patterns, by Country (Log Transformation)

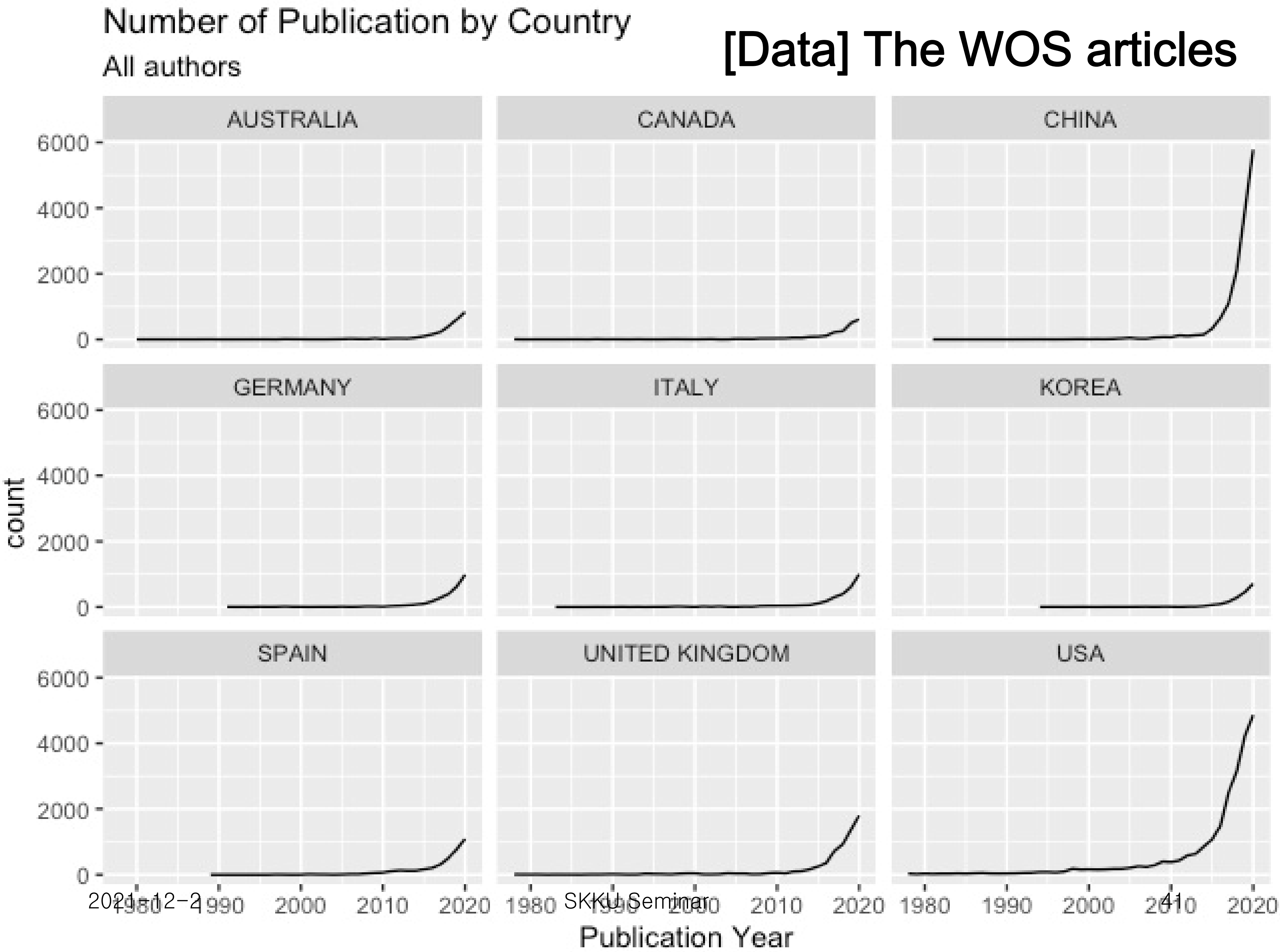
Growth in Numbers: Log Transformation



[Data] Distribution of the WOS articles

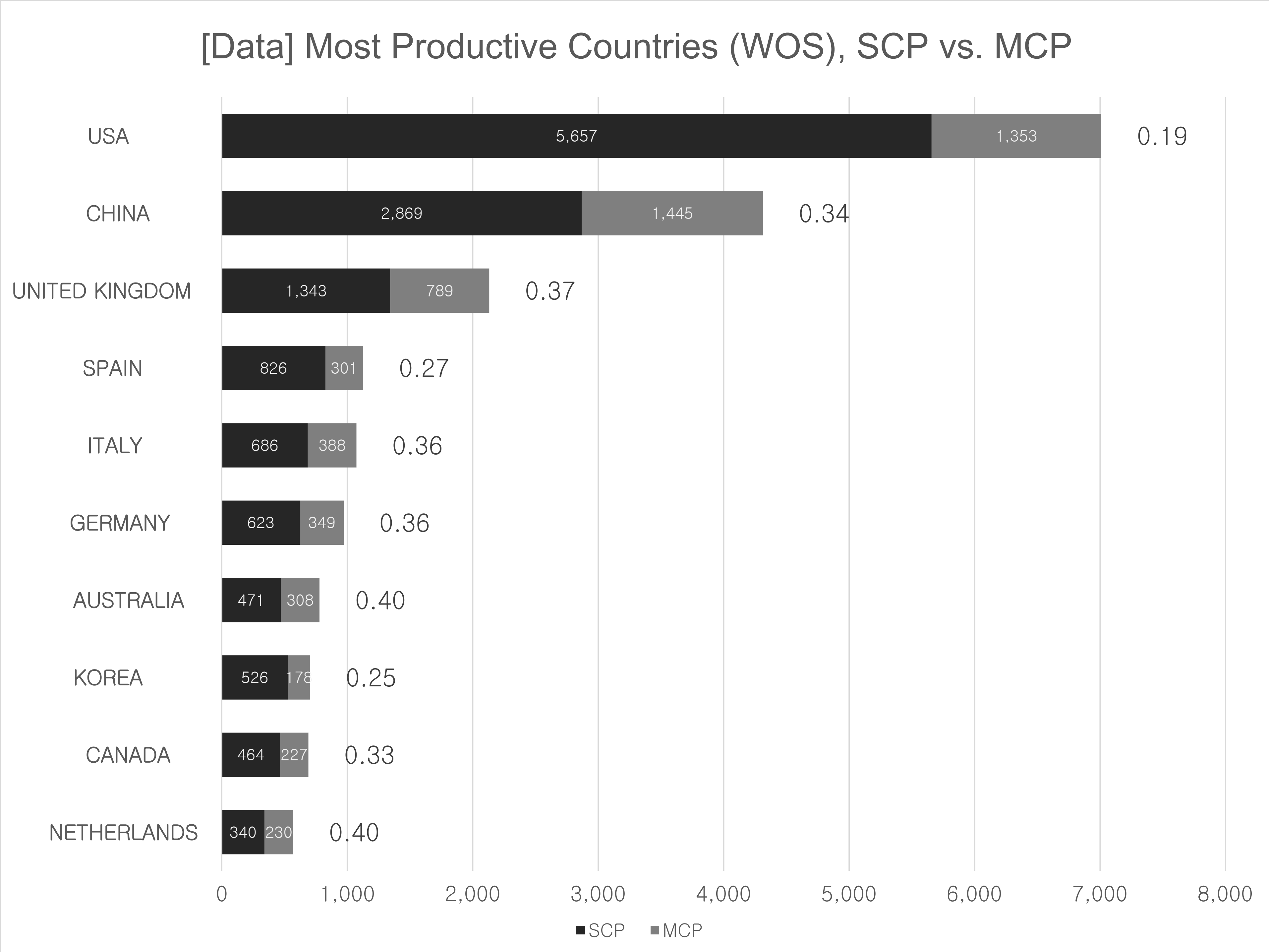
Corresponding Author's Countries						
	Country	Articles	Freq	SCP	MCP	MCP_Ratio
1	USA	7,010	0.27	5,657	1,353	0.19
2	CHINA	4,314	0.17	2,869	1,445	0.34
3	UNITED KINGDOM	2,132	0.08	1,343	789	0.37
4	SPAIN	1,127	0.04	826	301	0.27
5	ITALY	1,074	0.04	686	388	0.36
6	GERMANY	972	0.04	623	349	0.36
7	AUSTRALIA	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						

[Data] The WOS articles

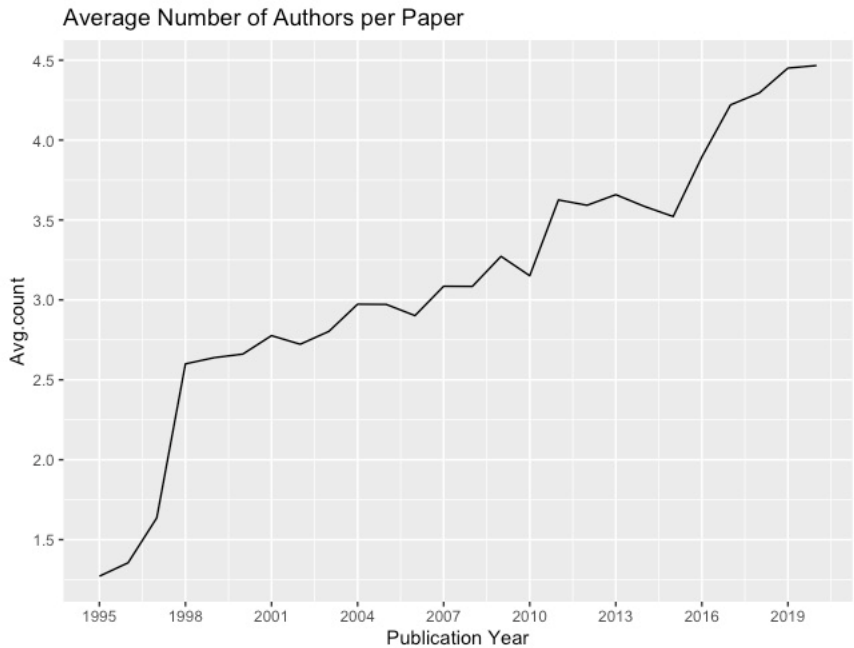


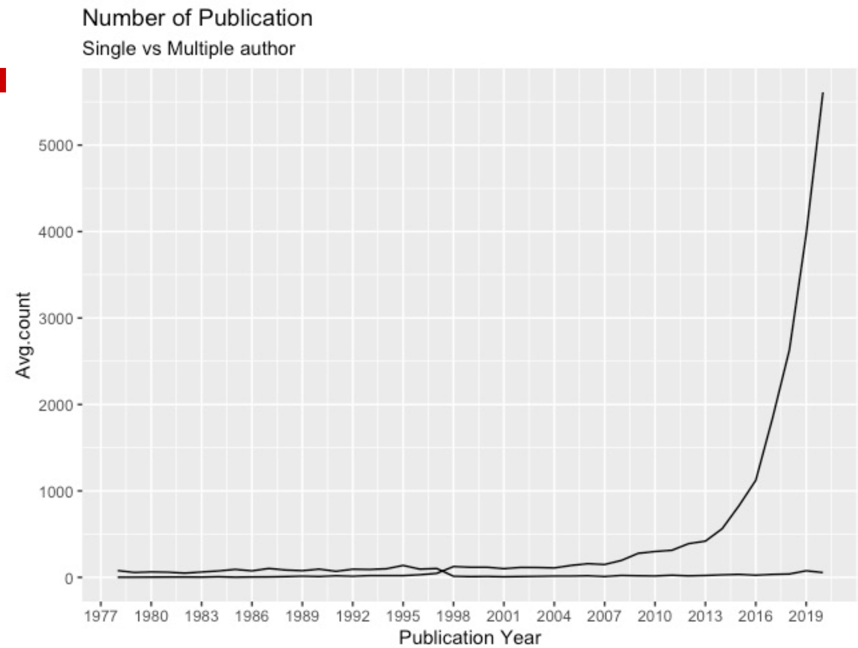
Observation (3): Rising trend of the multi-country collaboration and of the co-authorship

- Average number of authors & the multi-country research collaboration: Both rising continuously
  - About 80% of all WOS articles published in the U.S. journals being MCP (Multi Country Papers), with lower rates for other countries (China ~ 66%)
  - Average number of authors among the WOS articles: 2.5 in 1998 to 4.5 in 2020; Explosive growth of MCP articles from 2015 (same patterns observed from individual countries)
  - The international collaboration patterns by Korean researchers:
    - Mostly with the US authors (over 35%), followed by China, Germany, UK, Japan, & Australia
    - Pretty much no collaboration before 2006; Explosive growth since 2016

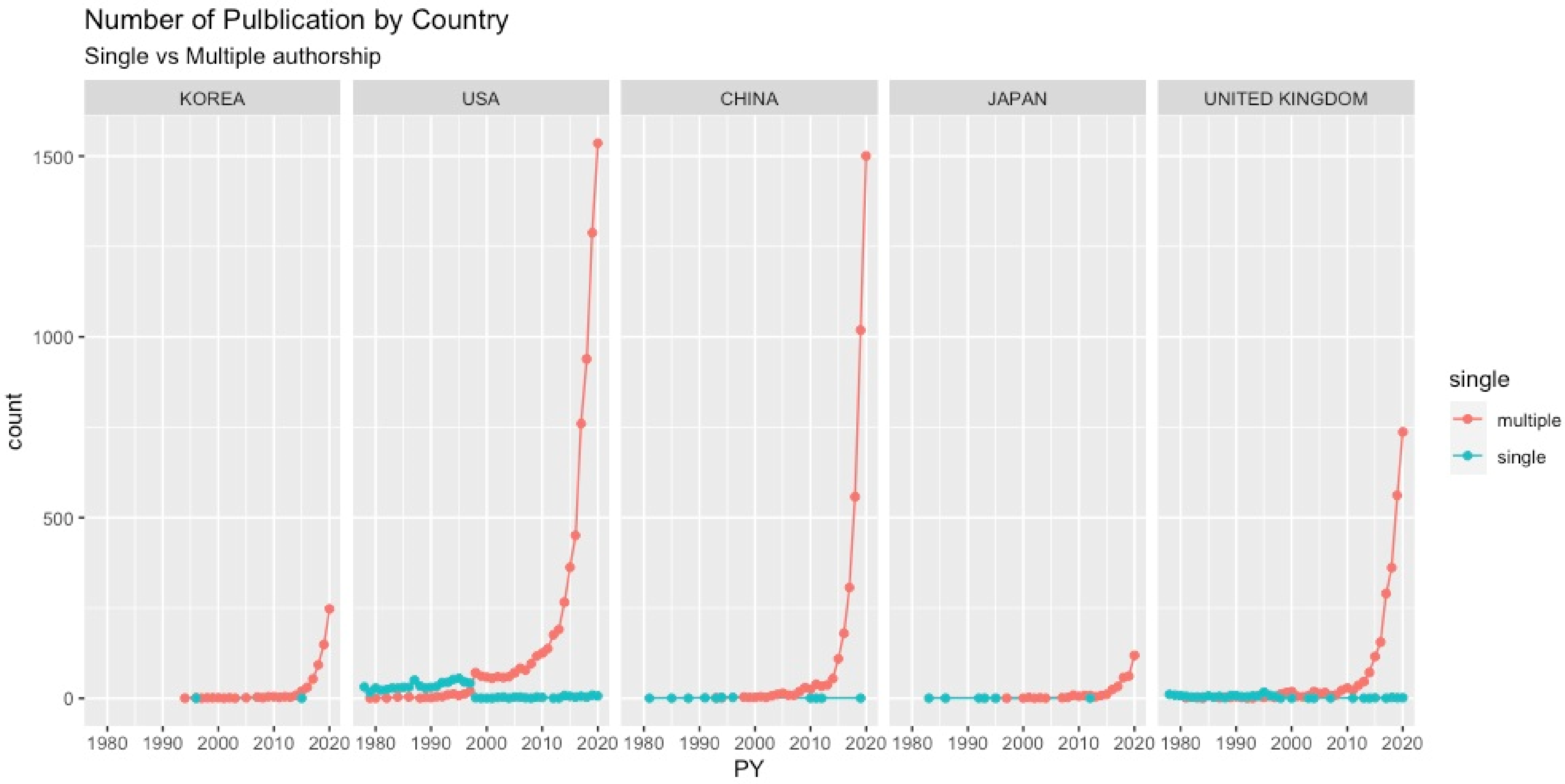

[Data] Most Productive Countries (WOS), SCP vs. MCP
USA
5,657
1,353
0.19
CHINA
2,869
1,445
0.34
UNITED KINGDOM
1,343
789
0.37
SPAIN
826
301
0.27
ITALY
686
388
0.36
GERMANY
623
349
0.36
AUSTRALIA
471
308
0.40
KOREA
526
178
0.25
CANADA
464
227
0.33
NETHERLANDS
340
230
0.40
0
1,000
2,000
3,000
4,000
5,000
6,000
7,000
8,000
SCP
MCP

[Data] Trends of co-authorship & cross-country collaborations

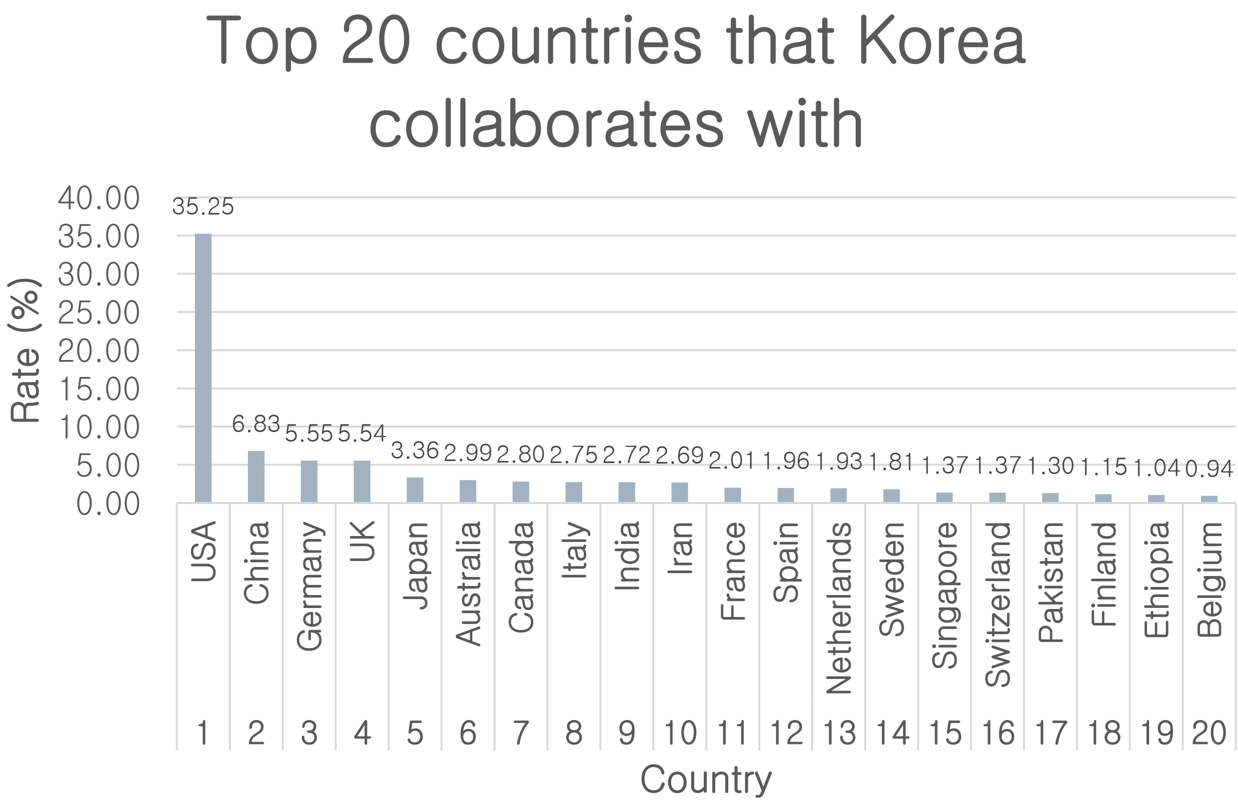




[Data] Distribution of the WOS articles



[Data] Collaboration Pattern (PubMed):
International Collaboration (Korea)



	Country	Count	Rate(%)
1	USA	4,271	35.25
2	China	827	6.83
3	Germany	673	5.55
4	UK	671	5.54
5	Japan	407	3.36
6	Australia	362	2.99
7	Canada	339	2.80
8	Italy	333	2.75
9	India	330	2.72
10	Iran	326	2.69
11	France	243	2.01
12	Spain	238	1.96
13	Netherlands	234	1.93
14	Sweden	219	1.81
15	Singapore	166	1.37
16	Switzerland	166	1.37
17	Pakistan	158	1.30
18	Finland	139	1.15
19	Ethiopia	126	1.04
20	Belgium	114	0.94

[Data] Collaboration Pattern (PubMed):

International Collaboration (Korea)

1991 ~ 1995		
1	South Korea	27

1996 ~ 2000		
1	South Korea	68

2001 ~ 2005		
1	South Korea	281

2006 ~ 2010		
1	South Korea	872
2	USA	3
3	Japan	2
4	Canada	1

2011 ~ 2015		
1	USA	389
2	China	109
3	UK	51
4	Japan	40
5	Germany	39
6	Italy	34
7	Spain	32
8	Australia	21
9	Canada	20
10	France	12
11	Belgium	11
12	Czech Republic	9
13	India	9
14	Malaysia	9
15	Pakistan	9
16	Netherlands	7
17	Taiwan	7
18	Austria	6
19	Greece	6
20	Poland	6
21	Singapore	6
22	Luxembourg	4
23	New Zealand	4
24	Portugal	4
25	Sweden	4

2016 ~ 2020		
1	USA	3,830
2	China	688
3	Germany	621
4	UK	618
5	Japan	354
6	Australia	341
7	Iran	324
8	Canada	323
9	India	308
10	Italy	278
11	France	230
12	Netherlands	227
13	Sweden	215
14	Spain	202
15	Switzerland	161
16	Singapore	154
17	Pakistan	149
18	Finland	132
19	Ethiopia	126
20	Belgium	102
21	Poland	102
22	Portugal	102
23	Brazil	100
24	Taiwan	82
25	Denmark	79

Observation (4): Variations across the countreis in the “quality” dimensions

- Indicators of the “quality” dimension of the articles
  - WOS: The number of citations per article being highest for the US articles (22.1), followed by UK (19.6) and Netherlands (15.2); But fairly low for Korea (9.8) and China (9.8)
  - PubMed: Leading countries in terms of the “newness” of the research topics – USA, France, Germany, and Japan (about 23 years’ average time lag for MeSH Heading keywords); Longer for Korea (25.7 years) and China (25.3 years)
  - Intertemporal trends of the top 10 PubMed MeSH headings (USA vs. Korea): Two surges for USA, only one surge for Korea (from 2016)
  - Impact of COVID19 on the PubMed MeSH headings (the last figure)

[Data] 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

[Data] Newness of the research topics (PubMed)

"Newness" of Topics by MeSH Age

Total MeSH Headings

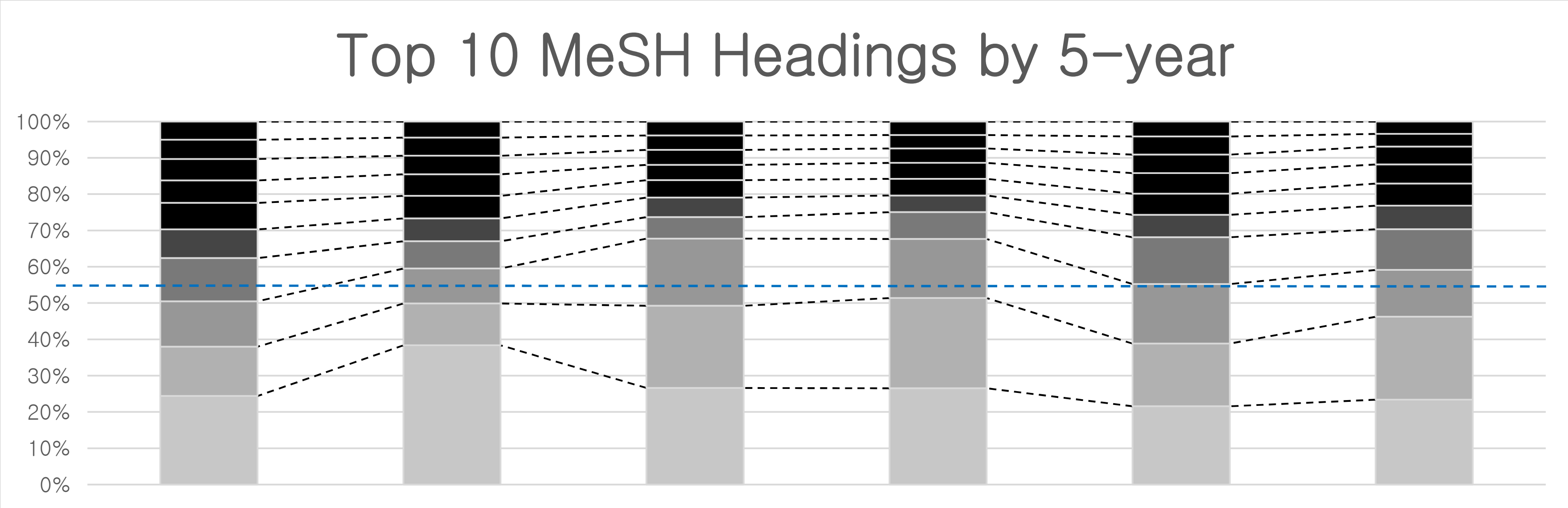
	Country	Sum. (MeSH Age)	Total Count (Article)	Avg. (MeSH Age)
1	Germany	4,067,471	155,373	26.2
2	France	1,854,449	70,681	26.2
3	USA	19,282,407	734,135	26.3
4	Japan	2,998,968	112,909	26.6
5	UK	4,113,648	152,001	27.1
6	Canada	2,069,033	75,999	27.2
7	Switzerland	971,324	35,758	27.2
8	Netherlands	1,507,057	54,966	27.4
9	Italy	2,373,947	85,842	27.7
10	Spain	1,413,584	49,963	28.3
11	Australia	1,525,376	52,812	28.9
12	India	1,147,221	39,614	29.0
13	China	7,153,615	245,107	29.2
14	South Korea	1,682,656	55,677	30.2
15	Brazil	1,292,673	42,256	30.6

Major topic of MeSH Headings

	Country	Sum. (MeSH Age)	Total Count (Article)	Avg. (MeSH Age)
1	USA	2,323,299	101,181	23.0
2	France	219,370	9,372	23.4
3	Germany	470,789	20,003	23.5
4	Japan	303,831	12,863	23.6
5	Italy	258,077	10,711	24.1
6	Switzerland	119,539	4,967	24.1
7	Canada	255,851	10,507	24.4
8	UK	529,376	21,623	24.5
9	Spain	184,541	7,450	24.8
10	Netherlands	169,766	6,806	24.9
11	India	137,596	5,459	25.2
12	China	881,420	34,838	25.3
13	South Korea	175,603	6,836	25.7
14	Australia	188,267	7,269	25.9
15	Brazil	139,553	5,060	27.6

[Data] Top 10 MeSH Headings & Their Weights (PubMed)

Transition of Digitalization Topic: MeSH Headings

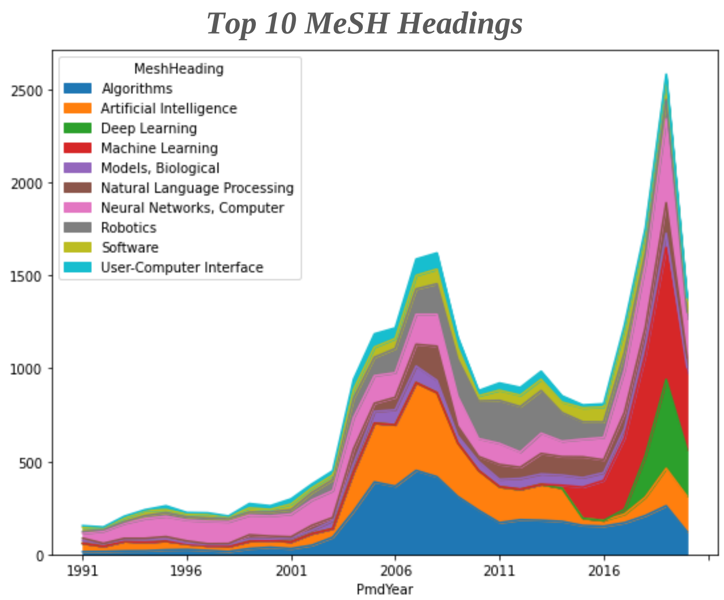


Rank	1991-1995	1996-2000	2001-2005	2006-2010	2011-2015	2016-2020
10	Algorithms	Radiographic Image Enhancement	Image Processing, Computer-Assisted	Natural Language Processing	Signal Processing, Computer-Assisted	Transcriptome
9	Software	Diagnosis, Computer-Assisted	Software	Software	Models, Biological	Models, Theoretical
8	Radiographic Image Enhancement	Robotics	Models, Neurological	User-Computer Interface	Natural Language Processing	Support Vector Machine
7	Computer Simulation	Signal Processing, Computer-Assisted	User-Computer Interface	Subtraction Technique	Models, Theoretical	Software
6	Signal Processing, Computer-Assisted	Computer Simulation	Models, Biological	Models, Biological	Software	Robotics
5	Diagnosis, Computer-Assisted	Expert Systems	Signal Processing, Computer-Assisted	Signal Processing, Computer-Assisted	Support Vector Machine	Artificial Intelligence
4	Image Processing, Computer-Assisted	Algorithms	Robotics	Robotics	Robotics	Deep Learning
3	Artificial Intelligence	Image Processing, Computer-Assisted	Artificial Intelligence	Neural Networks, Computer	Artificial Intelligence	Algorithms
2	Expert Systems	Artificial Intelligence	Algorithms	Artificial Intelligence	Neural Networks, Computer	Neural Networks, Computer
1	Neural Networks, Computer	Neural Networks, Computer	Neural Networks, Computer	Algorithms	Algorithms	Machine Learning

[Data] Top 10 MeSH Headings, Trends (USA)

Digital health research trends: USA (Cont.)

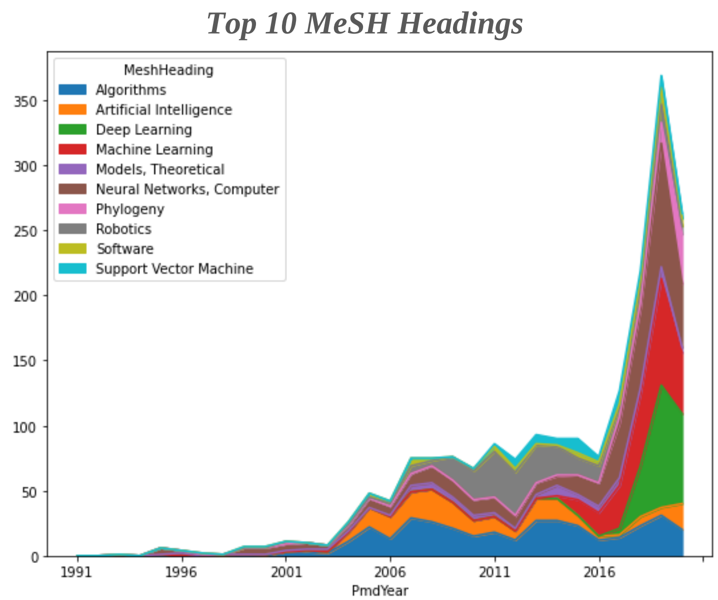
MeshHeading / PmdYear	Algorithms	Artificial Intelligence	Deep Learning	Machine Learning	Models, Theoretical	Neural Networks, Computer	Phylogeny	Robotics	Software	Support Vector Machine
1991	0	0	0	0	0	0	0	0	0	0
1993	0	0	0	0	0	1	0	0	0	0
1994	0	0	0	0	0	0	0	0	0	0
1995	1	0	0	0	0	5	0	0	0	0
1996	0	0	0	0	2	2	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



[Data] Top 10 MeSH Headings, Trends (Korea)

☐ Digital health research trends: Korea (Cont.)

MeshHeading / 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



[Data] Top 10 MeSH Headings, Impact of COVID19

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

Findings from the Meta Analyses: A Summary

- Outcomes from "the science of science" analyses:
  - An explosive growth of the DE-related articles from around 2015 (both for WOS & for PubMed)
  - "Quantity" of DE-research being dominated by three (US, China, & UK); Korea being reasonably high-ranked (8^{th} in both)
  - Rapid and continuous growth in co-authorship and multi-country research collaborations (for both DBs, MCP papers↑)
  - "Quality" of DE-research being dominated by a different group of countries (US, Germany, France, UK, Japan); Korea and China being lagged in this dimension (for WOS & PubMed)

Policy Implications
(& Policy Tasks, PT)

The DE-Related Policy Initiatives in Korea

- Various DE-related Policy Initiatives being implemented in Korea during the last several years
  - Three data legislations (19), for use & protection of personal data
  - Data Exchanges (20), MyData (21), OpenData (on-going)
  - 'Digital New Deal' (20 for V-1.0 and 21 for V-2.0)
  - SmartCity, SmartFarm, and other Smart X initiatives (on-going)

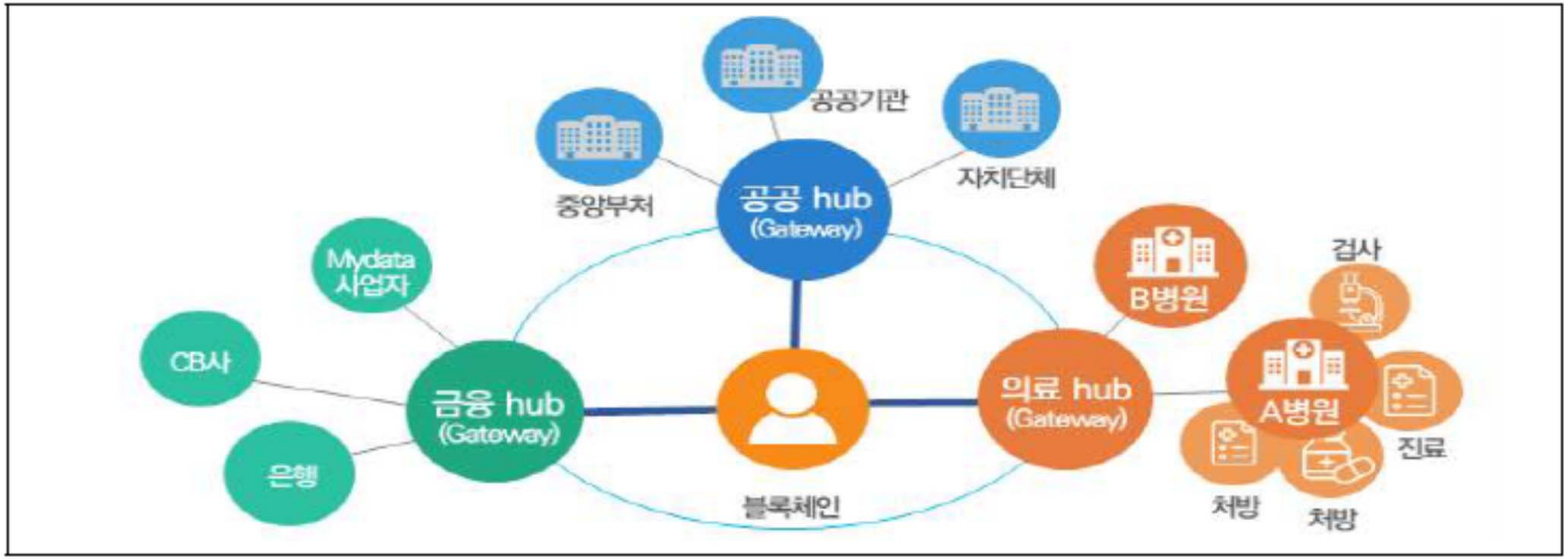
- What do the meta analyses performed imply in initiating those policies? ➔ Several policy tasks to be discussed subsequently

PT (1): A cross-sectoral & interdisciplinary approach needed (in Korea)

- In promoting the DE-ecosystem in Korea, a cross-sectoral approach being needed
  - A big sectoral difference being observed; “Finance” being the most active sector in fostering DE-ecosystem (w/ Data Exchanges and MyData having implemented) ➔ A cross-sectoral policy (e.g., anonymication, API, analytics, & other standards setting) being needed
  - Categorizing digital data, and setting a customized strategy (e.g., (1) public data to share with everyone, (2) common data to share within a network, and (3) private data to exchange with price)
  - A more result-oriented DE policies to be initiated (e.g., the “All of US” project in the US for medical research)

[Data] The Multi-Sectoral 'MyData' Initiatives in Korea

❖ 분야별(금융, 의료 등) **허브(hub, 관문)를 통한** 본인정보 제공·유통 및 **상호연계**



Source: MSA (2020)

All of Us RESEARCH PROGRAM

NIH National Institutes of Health
Turning Discovery Into Health

https://allofus.nih.gov

The Precision Medicine Initiative Cohort Program – Building a Research Foundation for 21st Century Medicine

출처: 강건욱 (2020)

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.

출처: 강건욱 (2020)

N Engl J Med 2019; 381:668-676

PT (2): An interdisciplinary research and PPAP to be promoted

- For the purpose of developing an effective DE-ecosystem, an interdisciplinary research and PPAP (Public·Private·Academic·Partnership) to be promoted
  - A good supply of well-trained specialists being essential in promoting DE-ecosystem ➔ Develop interdisciplinary data science programs (schools or training programs) to create regional innovation clusters
  - Integration of academia with public & private sectors to enhance the socio-economic effects of DE-ecosystem (e.g., "Bristol is Open" in UK as an example PPAP case; and "EU Replicate" programs for the regional SmartCity initiatives)

PT (3): Promoting data- & evidence-based decision makings

- In all three sectors, promote data- & evidence-based decision makings by utilizing DE-ecosystem
  - Finance: A more efficient & real time credit risk management (by using alternative data, AI-enabled analytics, and so on) ➔ A more accurate consumer segmentation and expansion of financial inclusion (e.g., to marginal borrowers, “thin filers”)
  - Real estate: A more transparent RE market data (e.g., RE price indices); Reducing search and information cost via RE platforms (the political economy issues to be addressed, more active role by the government in promoting DE)
  - Medical service: A more active utilization of one’s own medical data (& linking that to MyHealthway); Via medical data and new analytics, promoting more innovations in medical services

PT (4): Promoting and regulating the new inter-sectoral business

- Defining the new industrial boundaries (e.g., conventional banks vs. BigTechs)
  - The traditional industrial boundaries being blurred more and more (Rising roles of BigTechs such as Google," Amazon, Ant Financial, PingAn Insurance, and Kakao)
  - More efficient and convenient services provided by the Smart X industries (e.g., FinTech firms – "Banking is necessary but banks are not!" & "Is bank still special?")
  - Need to have a LT development strategy for existing service providers and DPA-enabled Smart X service providers as to how to regulate and promote their mutual welfare gains

Alibaba group Alibaba.com

Affiliated companies	Description & key business area
Alibaba.com	The leading wholesale marketplace for global trade
1688.com	The leading integrated domestic wholesale market place in China
Alibaba Cloud	A cloud computing service provider
AliExpress	A global retail e-commerce platform
Alimama	A marketing technology platform
Taobao.com	The China's largest mobile commerce platform
TMALL.com	The China's largest B2C platform
Cai Niao	A logistics data platform operator
Ant Finanial	A technology company offering inclusive financial services

Ant Financial Group

Affiliated companies	Description & key business area
AliPay	A mobile payment platform with 520m+ users, and business partners across over 15 countries
Yu'e Bao	The largest money market fund in the world, managing $221b
Ant Fortune	MPL for Ant Financial and third-party financial products, with 180m users.
Ant Insurance Service	An insurance service firm with 400m users, offering its own and 80+ insurance companies' products.
Zhima (Sesame) Credit	A credit scoring company, using social networks and payments history with about 260m users.
Ant Cash Now	A credit company for quick funding for AliPay users, based on user risk profiles.
Ant Credit Pay	A consumer lending company with 100m active users, having lent $95b to consumers through Q1'17.

PT (5): Linking DE-ecosystem to Smart X (for job creation and regional development)

- Increasing the social effects of DE-ecosystem by explicitly linking it to Smart X initiaties
  - For both job creation and regional development, link local DE-ecosystem to SmartCity, SmartFarm, SmarkMedicalService, and other Smart X initiatives
  - Promote DE related policies to enhance area-specific regional development policies, by benchmarking the good PPAP cases as well as by explicitly relating them to local job creation (for balanced regional development)
  - Initiating more "result-oriented" DE policies (e.g., job creation); And monitoring their progresses periodically (e.g., measuring social BC for DE- and Smart X-related policy initiatives)

Concluding Remarks

Concluding Remarks

- What's been done?
  - A tri-sectoral study to gauge the current states in developing a DE-ecosystem along with policy tasks needed
  - Findings from the literature review and the meta analyses being shared; Policy implications & tasks also being discussed
- What's to be done (i.e., future research agenda)?
  - Developing KPIs (e.g., Indices for policy uncertainty, & digitaliz.)
  - Analyzing & strategizing local DE-ecosystem (including a more meaningful classification system for local economies)
  - Relating DE-ecosystem to local housing & labor market outcomes

Thank you!