RIETI Discussion Paper Series 19-E-022

Value of Data:
There's No Such Thing as a Free Lunch in the Digital Economy

LI, Wendy C.Y.
U.S. Bureau of Economic Analysis

NIREI, Makoto
RIETI

YAMANA, Kazufumi
Kanagawa University

Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 19-E-022
March 2019

Value of Data: There's No Such Thing as a Free Lunch in the Digital Economy

Wendy C.Y. Li (U.S. Bureau of Economic Analysis)
Makoto Nirei (University of Tokyo and RIETI)
Kazufumi Yamana (Kanagawa University)

Abstract

The Facebook-Cambridge Analytica data scandal demonstrates that there is no such thing as a free lunch in the digital world. Online platform companies exchange "free" digital goods and services for consumer data, reaping potentially significant economic benefits by monetizing data. The proliferation of "free" digital goods and services pose challenges not only to policymakers who generally rely on prices to indicate a good's value but also to corporate managers and investors who need to know how to value data, a key input of digital goods and services. In this research, we first examine the data activities for seven major types of online platforms based on the underlying business models. We show how online platform companies take steps to create the value of data, and present the data value chain to show the value-added activities involved in each step. We find that online platform companies can vary in the degree of vertical integration in the data value chain, and the variation can determine how they monetize their data and how much economic benefit they can capture. Unlike R&D that may depreciate due to obsolescence, data can produce new values through data fusion, a unique feature that creates unprecedented challenges in measurements. Our initial estimates indicate that data can have enormous value. Online platform companies can capture the most benefit from the data because they create the value of the data and because consumers lack knowledge regarding the value of their own data. As trends such as 5G and the Internet of Things are accelerating the accumulation speed of data types and volume, the valuation of data will have important policy implications for investment, trade, and growth.

Keywords: Artificial Intelligence, Data, Data Monetization, Data-driven Business Model, Intangible Capital, Innovation, Online Platform
JEL Classification: O3

This study is conducted as a part of the project "Research on New Technologies, Economic Growth, and Industrial Structure" undertaken at Research Institute of Economy, Trade and Industry (RIETI). The views expressed are those of the authors and do not necessarily reflect those of the U.S. Department of Commerce or the Bureau of Economic Analysis. We thank Daniel Levinthal, Masayuki Morikawa, Sadao Nagaoka, Dylan Rassier, Paul Roberts, Gabriel Quiros Romero, Rahul Telang, and participants in the *2018 IP Statistics for Decision Makers Conference* and *the Sixth IMF Statistical Forum* for helpful comments.

1. Introduction

Because of improved programming capabilities and the rapid price decline of information technology hardware and services, new business models have emerged and many of them are embodied in online platforms. For example, online sharing platforms like Uber increase the efficiency of underutilized assets and lower the consumption prices of the services. E-commerce platforms, such as Amazon Marketplace, have greatly reduced transaction costs for many small and medium sized enterprises (SMEs) to sell products across states and borders. Online platforms, mostly created and run by young companies, are physical-asset-light but have grown fast and deeply disrupted many industries (Li, Nirei and Yamana, 2017, 2018). A prominent example is Airbnb, a company that has only 1.7% of the employee size of Marriott International, but more listed properties than the top five global hotel groups combined (Hartmans, 2017). Moreover, online platforms have been growing rapidly in scale. For example, based on Census data, Hathaway and Muro (2017) show that the U.S. ridesharing service has been experiencing a hyper-growth rate and can take over the taxi services in the near future.

Most online platforms have been providing digital goods and services to consumers at seemingly zero monetary cost, and economists have been attempting to measure the welfare effects related to "free" digital goods and services. For example, Brynjolfsson et al. (2018) estimate that Facebook creates US $21.4 billion consumer surplus in 2017 in the U.S. alone. However, the Facebook-Cambridge Analytica data scandal demonstrates that there is no such thing as a free lunch in the digital world. In fact, consumers exchange their personal data for "free" digital goods and services. As large data holders, online platform companies like Google and Facebook can reap potentially significant economic benefits by providing data targeting services and/or licensing the use of the data to third parties. Therefore, phrases such as "free goods" are misnomers. Welfare analysis on digital goods and services without considering the value of data can mislead policy analysis.

Online platform companies are physical-asset-light but can be extremely profitable. They have collected copious amounts of rich data through their online platforms, monetized the data, and created vast amounts of value from data. For example, Booking Holdings, the world's leading online travel platform company, reported a gross profit margin of 98% in 2017 and of 95% averaged over three years (SEC, 2017). At its Amsterdam headquarters, 90% of Booking's employees are engineers (Yin, 2018). While being a data company, Booking outsources its data centers to take advantage of cheap cloud services. Another example is Facebook: when it went public in 2011, the value of its total assets was reported at US $6.3 billion, but its market valuation reached as high as US $104 billion (SEC, 2012). The huge gap between the two numbers implies the enormous value of its intangible assets, including the value of data. Facebook exchanges free social media services for user data, and conducts analytics on user data to provide third parties with data targeting services, currently mainly data targeted advertising. In 2017, its advertising revenue was US $39.9 billion, contributing to 40% of its annual sales growth (Forbes, 2017).

Data are crucial for AI revolution and firms' competitiveness, but they are intangible capital whose value is very difficult to measure. On the one hand, data are not tangible capital that suffers wear and tear. On the other hand, data are not regular intangibles like R&D capital that may depreciate due to obsolescence (Li and Hall, 2018). The aggregation and recombination of data can create new value. Furthermore, it is well known that getting data and information from online platform companies is difficult (Demunter, 2018). These unique features of data pose challenges to valuing data.

Nevertheless, what gets measured gets managed. Two examples can help us visualize the size of the value of data. The first example is Apple. By charging app developers 30% commission of their sales for accessing Apple's consumer data, Apple has earned US $42.8 billion in revenue in the past decade (Frier, 2018). The second example is ITA Software versus Farecast. ITA Software is a large airline reservation network collecting the detailed transaction data of U.S. airline tickets. When Farecast was an independent company, it purchased data from ITA Software and conducted analytics to predict airfares (Mayer-Schönberger and Cukier, 2014). Farecast was acquired by Microsoft in 2006 for US $110 million. However, ITA Software, the data owner, was acquired by Google two years later for US $700 million. The acquisition price difference between the two firms implies that data can potentially be more valuable than analytics capabilities. In the age of AI implementation, as AI is becoming cheap, data are emerging as the core to govern the overall power and accuracy of an algorithm (Agrawal et al., 2018; Beck and Libert, 2019; Lee, 2018). Moreover, how firms utilize their data analytics to monetize data relies on their business models. When Google purchased ITA Software, it might already have a business plan to monetize the data. In 2011, three years after the purchase of ITA Software, Google launched Google Flights, which has become the most popular flight search online platform in the U.S. (Whitmore, 2018).

The substantial market valuation of data shown in the ITA Software-Farecast-Google Flights example highlights the importance of measuring data activities related to online platforms. The measurement of the value of data can provide important information not only for public policies such as digital trade and national data policies, but also for corporate strategies such as investment and outsourcing decisions in data and data-driven decision making processes. Moreover, this kind of information is also important for investors to understand firm fundamentals and facilitates capital flows to innovative firms in the era of data-driven economy.

Online platforms can differ in their underlying business models. Business model represents how a firm creates and delivers value for its customers while also captures value for itself in a repeatable way (Johnson, 2018). Online platform companies are data companies, and their underlying business models determine what type of data they collect, how data flow within online platform networks, how the companies monetize the data, and what consumers can gain by exchanging their data. Therefore, it is necessary to examine the value creation in different types of online platforms to understand the common characteristics or possible variations.

In this paper, we conduct case studies to analyze data activities in seven major types of online platforms classified by the Organization for Economic Co-operation and Development (OECD, 2018a). We show how online platform companies take steps to create the value of data, and present a data value chain to show the value-added data activities involved in each step. We also present a physical supply chain of data monetization to illustrate what investment and outsourcing options the companies face at each stage. We find that online platform companies can vary in the degree of vertical integration in the data value chain, and the variation can determine how they monetize their data and how much economic benefits they can capture. Our initial estimates show that the value of data is enormous and depends crucially on online platform companies' data-driven business models. Moreover, online platform companies can capture most benefits of the data, because they create the value of data and because consumers lack knowledge to value their own data.

2. Online Platforms: Major Types and Data Activities

2.1 Typologies of Online Platforms

In this research, we adopt the OECD definition that an online platform is "digital services that facilitate interactions between two or more distinct but interdependent sets of users (whether firms or individuals) who interact through the service via the internet" (OECD, 2018a). For example, based on this definition, Amazon Marketplace is an online platform, but Amazon direct sales is not (OECD, 2018b).

In addition to some studies attempting to classify online platforms (Chen et al., 2018; Demunter, 2018; van de Ven, 2018), OECD conducted a multiple-year project on online platforms and has identified several typologies (OECD 2018a). As indicated in this official document that has been reviewed by OECD countries, there are many typologies of online platforms, and the choice among them depends on the research or practice need at hand. In some situations, it may be useful to apply several topologies at once. Since online platforms continue to evolve in different industry sectors across the globe, no single typology can cover all online platforms. Moreover, some typologies use the same name for a certain type of online platforms, but the included companies differ. Therefore, one should pay attention to the companies involved when comparing the typologies from different studies.

We select the following seven major types of online platforms identified by OECD to study the associated data activities and the data monetization strategies:

Type I: E-commerce Platform

Type II: Online Sharing Platform

Type III: Fintech Platform

Type IV: Online Social Network Service Platform

Type V: Online Matchmaking Platform

Type VI: Online Crowdsourcing Platform

Type VII: Online Search Platform

These types of online platforms have different underlying business models. For each type of online platform, we conduct a case study to examine its underlying data-driven business model: data flow, value creation for consumers, value creation for third parties, and how an online platform company monetizes its data. Due to the serious limitation in publicly accessible information, we focus on the companies for which some public data or reports are available.

2.2 Type I: E-commerce Platform

Type I is the e-commerce platform, and our case study is Amazon Marketplace (Figure 1). Amazon Marketplace is an online platform that facilitates sales between consumers and third-party sellers. On the one hand, it offers consumers a place to purchase a wide range of products from more selections with cheaper prices. On the other hand, it allows third-party sellers to access one of the world's largest e-commerce markets in a cost-effective and time-efficient way.

Amazon charges third-party sellers a commission of approximately 30% of their sales (Mims, 2018). The commission pays for not only the cost of accessing one of the world's largest e-commerce markets but also the cost of "basic" access to Amazon's consumer data. For example, when a consumer purchases a good by cash in an offline supermarket, the supermarket and the third-party seller that offers the good do not obtain data about the consumer. However, if the customer pays by a credit or debit card, the supermarket but not the third-party seller will have some data about the consumer. By contrast, when a consumer purchases a good online through Amazon Marketplace, not only Amazon but also the third-party seller can acquire the consumer data. Nonetheless, there is a difference in the degree and the details of the data. The third-party seller can get the data displayed in the transaction; however, Amazon can obtain consumer data beyond the transaction data, including browsing history and clickstreams. Moreover, Amazon has all transaction data related to third-party sellers.

Figure 1: *Type I: E-commerce Platform*

Case Study: Amazon Marketplace

In terms of data flow, Amazon collects data on clickstreams, purchases, reviews, and locations from consumers.[3] Then, it conducts data analytics to provide data-targeting services to third-party sellers. For example, based on the geolocation data of consumers and demand forecast, it can provide third-party sellers with logistics consulting services such as where to build the warehouse. Bond (2018) reports that Amazon offers corporate clients premium data services, which include demand and trend forecasts, and the price for such premium data services starts from US $100,000 per year. In addition, Amazon gathers information of consumer price sensitivity

[3] Note that online platform companies can also collect data from third-party sellers such as where they ship the products if they choose to fulfill the orders by themselves. When online platform companies provide data targeting services, they can incorporate the profile of their third-party sellers.

by funding discounts on third-party products. Combining this price sensitivity data and other data, Amazon can conduct detailed profiling of each consumer and provide data-driven pricing strategy services to third-party sellers.

In 2017, e-commerce accounts for 10% of U.S. retail sales, and Amazon has 43% of the U.S. e-commerce market share (Molla, 2017). In addition, 50.5% of its e-commerce sales are conducted through third-party sellers on Amazon Marketplace (Statista, 2018). Based on the 2017 Amazon Marketplace's sales, US $139.5 billion, and the 30% commission charge to third-party sellers, Amazon's estimated annual revenue from the commission is US $41.8 billion (Amazon 10K report). While growing fast, Amazon's data targeted advertising revenue in 2017 amounted only to US $3 billion, merely 2.2% of its total revenue in that year. Compared to Facebook and Google, Amazon does not rely on advertising revenue.

2.3 Type II: Online Sharing Platform

Type II is the online sharing platform, and our case study is Booking.com (Figure 2). Booking.com is a leading online travel sharing platform that facilitates sales between consumers and property owners. On the one hand, it offers consumers a place to reserve rooms from many properties with discounted rates. On the other hand, it allows hotels or property owners to access one of the world's largest online travel markets and to reduce the inventory of their highly perishable goods or monetize their underutilized private rooms. Booking.com charges a 15% commission of the sales revenue from third-party sellers.

In terms of data flow, Booking.com collects data on clickstreams, purchases, reviews, and locations from consumers. It also conducts data analytics to provide third-party sellers with data targeting services, such as pricing strategy, demand forecast, and consulting services. It was reported that Booking.com's data analytics service on pricing strategy on average increased third-party sellers' sales revenue by 7% (Yin, 2018). The total number of its listed available private rooms is larger than that of Airbnb. Since Booking.com charges third-party sellers a 15% commission on their sales revenue, the 2017 estimated revenue from commissions alone is US $11.8 billion. At its Amsterdam headquarters, Booking has 1,800 engineers that account for 90% of its employees. The company outsources its data centers and benefits from cheap cloud services, another business strategy that makes it physical-asset-light but extremely profitable.

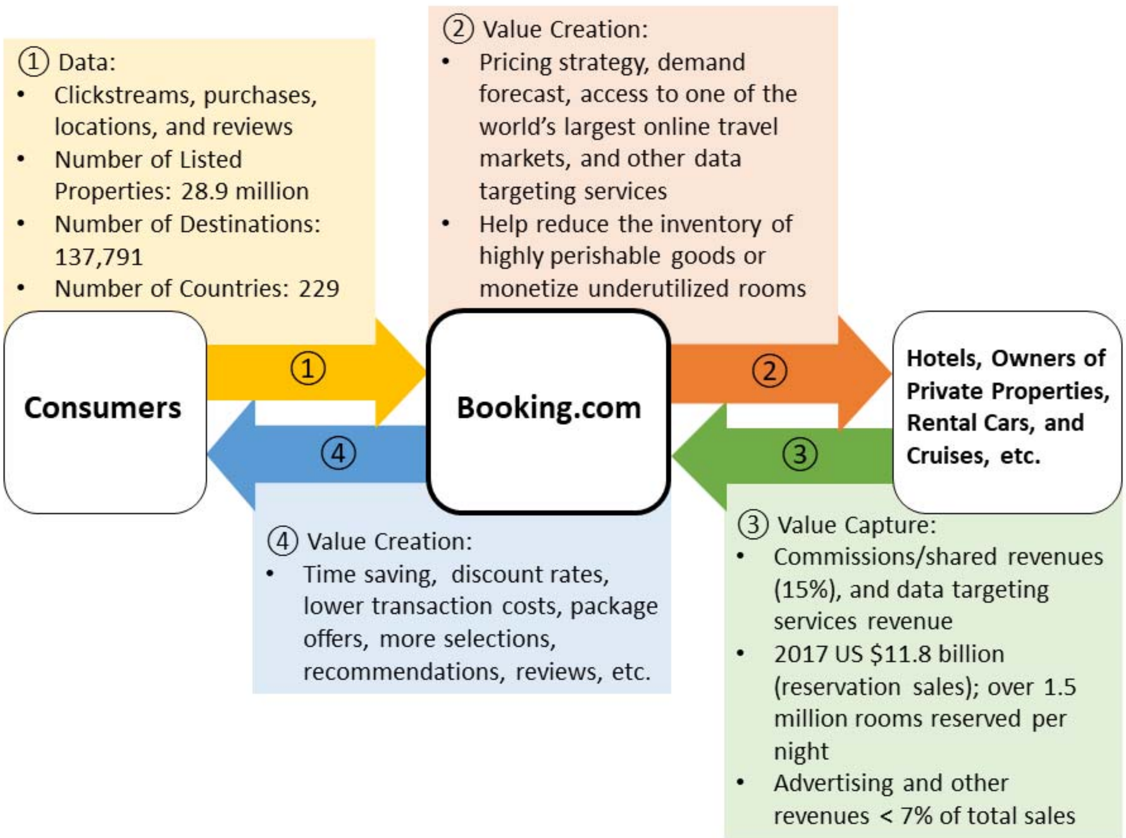

Figure 2: *Type II: Online Sharing Platform*

Case Study: Booking.com

What is the difference between Booking.com and Marriott International, the world's largest hotel chain and a middleman in the hotel industry? In the early 1980s, Marriott invented a business model by licensing its franchise and providing management services to real estate developers who own hotel properties. However, Booking.com's online platform and business model have allowed it to reach a much broader range of property owners and to have a greater scalability within and across regions. With more listed properties and a greater scalability, Booking.com can better collect consumer data than Marriott in terms of volume, type, and speed. Moreover, Marriott's 2017 gross profit margin, 16%, is far less than the 98% of Booking Holdings.

2.4 Type III: Fintech Platform

Type III is the Fintech platform, and our case study is Ant Financial (Figure 3). Ant Financial is China's largest online financial platform that facilitates financial transactions among financial institutions, merchants, and consumers. On the one hand, it offers consumers and microbusinesses a way to access the credit that was previously unavailable. On the other hand, it allows financial institutions to reach customers who previously have no credit history. To date, there are 870 million active users globally and the majority of them are in China.

In terms of data flow, Ant Financial collects data on clickstreams, daily consumption and lending behaviors, locations, and bank account information from consumers and microbusinesses. It conducts data analytics to provide data targeting services to corporate clients, such as credit scoring services to financial institutions and demand forecast to hotels. Currently, its third-party institutions include more than 200 banks, 60 insurance companies, and 700,000 stores. The reported revenue from Alipay, its online payment platform, is US $1 billion in 2017.

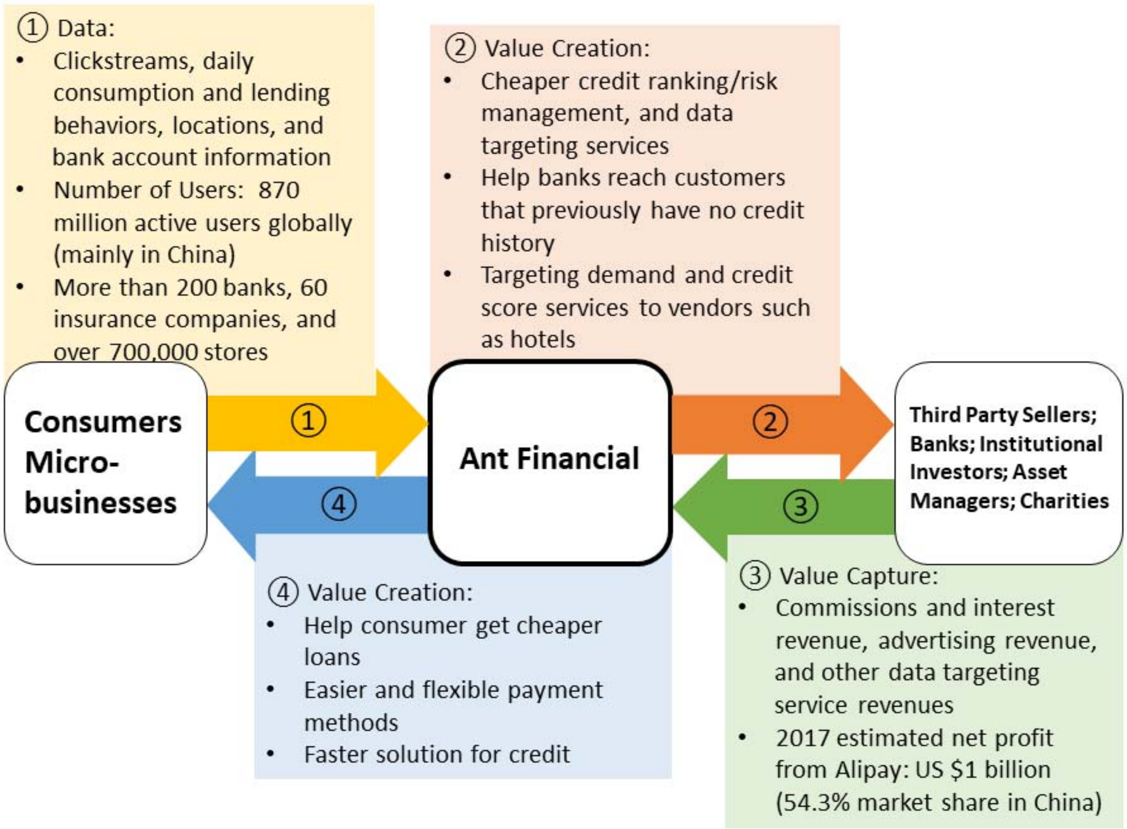

Figure 3: *Type III: Fintech Platform*

Case Study: Ant Financial

2.5 Type IV: Online Social Network Service Platform

Type IV is the online social network service platform, and our case study is LinkedIn (Figure 4). LinkedIn is a leading business and employment-oriented service platform that facilitates professional networking. On the one hand, it allows individuals to post their résumés and connect with professional friends. The professional network may also facilitate job search. On the other hand, it allows employers to post jobs and search potential candidates. To date, there are approximately 500 million users in over 200 countries.

In terms of data flow, LinkedIn collects data on clickstreams, work experience, qualifications, professional networks, work preference, and views from its members. LinkedIn then sells access to its member data to recruiters and sales professionals. Before it was acquired by Microsoft, its revenue came mostly from selling access to its member data. LinkedIn's revenue in 2015 was US $2.99 billion and in 2016, Microsoft purchased LinkedIn for US $26.4 billion.

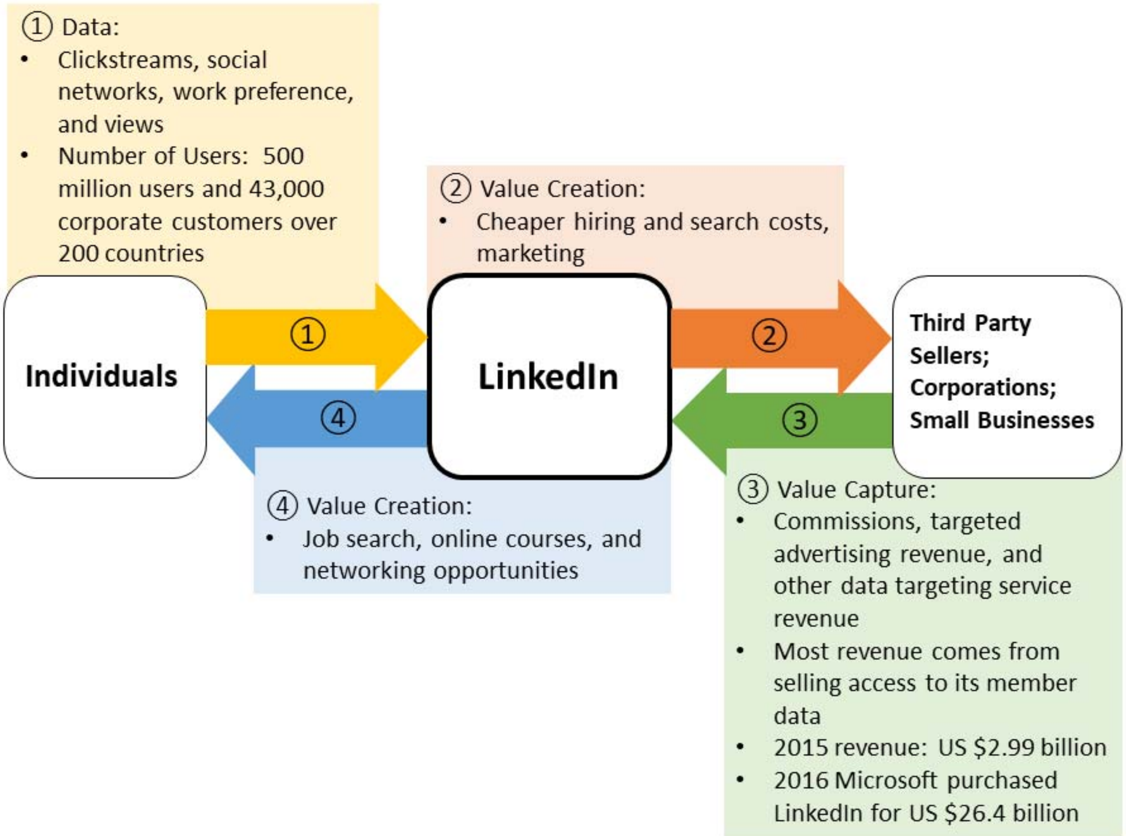

Figure 4: *Type IV: Online Social Network Service Platform*

Case Study: LinkedIn

2.6 Type V: Online Matchmaking Platform

Type V is the online matchmaking platform, and our case study is eBay (Figure 5). Some characteristics of eBay are similar to those in the type of e-commerce platform, but OECD classifies the company as a matchmaking online platform. eBay is a leading online auction platform that facilitates consumer-to-consumer, business-to-consumer, and business-to-business sales. It is free for buyers to use, but sellers are charged fees for listing items (after a limited number of free listings) as well as the sale. On the one hand, it provides the buyer with a convenient and cheaper way to purchase products and special collection items. On the other hand, it allows sellers to access a large online auction demand market. To date, there are 175 million active users in over 30 countries.

In terms of the data flow, eBay collects data on clickstreams, bidding histories, and payment histories from users. It then conducts data analytics to sell data targeting services. Employing approximately 5,000 data analysts, eBay reportedly has already experienced significant business successes through its data analytics (Ovenden, 2016).

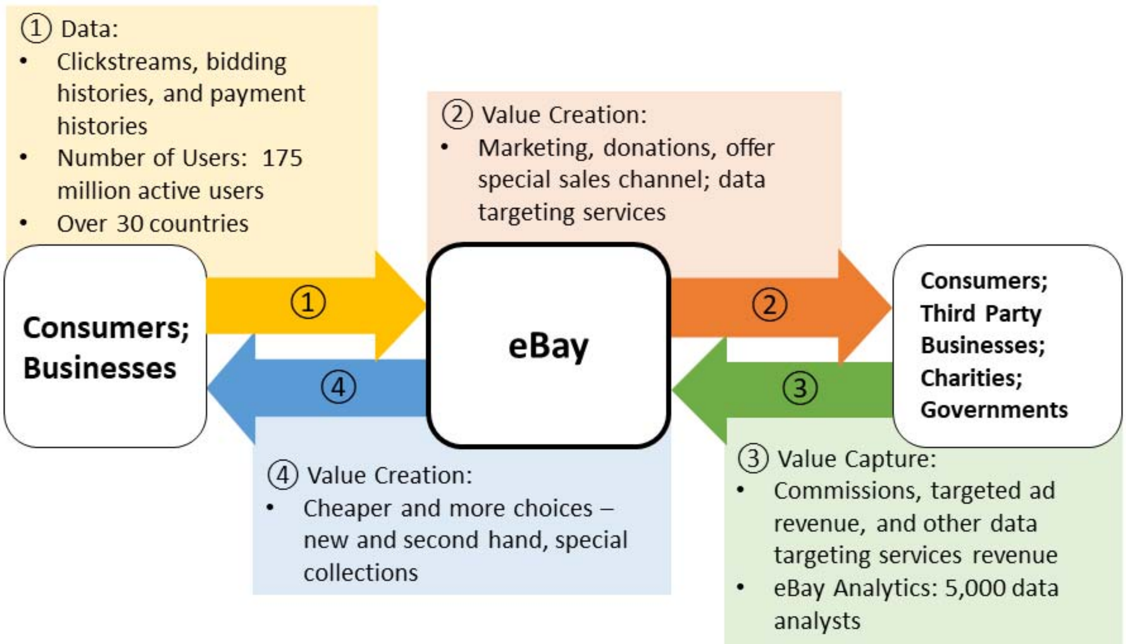

Figure 5: *Type V: Online Matchmaking Platform*

Case Study: eBay

2.7 Type VI: Online Crowdsourcing Platform

Type VI is the online crowdsourcing platform, and our case study is Waze (Figure 6). Waze is a popular crowdsourcing platform that facilitates data sharing among drivers. Drivers report accidents, traffic jams, speed and other information about road conditions. It provides drivers with real-time traffic updates, routing, nearby cheapest fuel prices, and other location-specific alerts.

In terms of the data flow, Waze collects data on map data, travel times, traffic information, and locations from drivers. It then conducts data analytics to provide data targeting services. For example, Waze can use data on traffic flow to provide a pricing strategy service to billboard owners. In 2013, Google bought Waze for US $1.3 billion to add social data to its mapping business (Cohan, 2013).

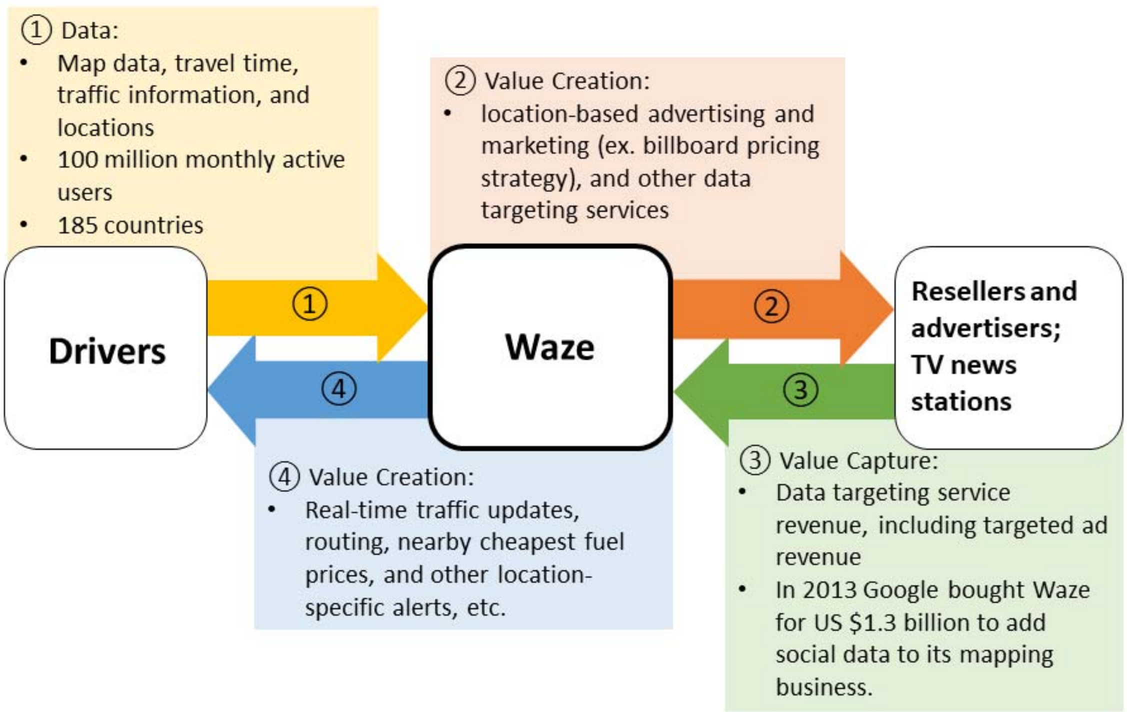

Figure 6: *Type VI: Online Crowdsourcing Platform*

Case Study: Waze

2.8 Type VII: Online Search Platform

Type VII is the online search platform, and our case study is Google Search (Figure 7). Google Search is currently the most popular online platform in the world. On the one hand, it provides individuals with a free, convenient, and relevant way to get information instantly. On the other hand, it allows advertisers and content providers to reach one of the world's largest user bases in an effective fashion. It also allows content providers to add search functionality to their web pages and monetize their content.

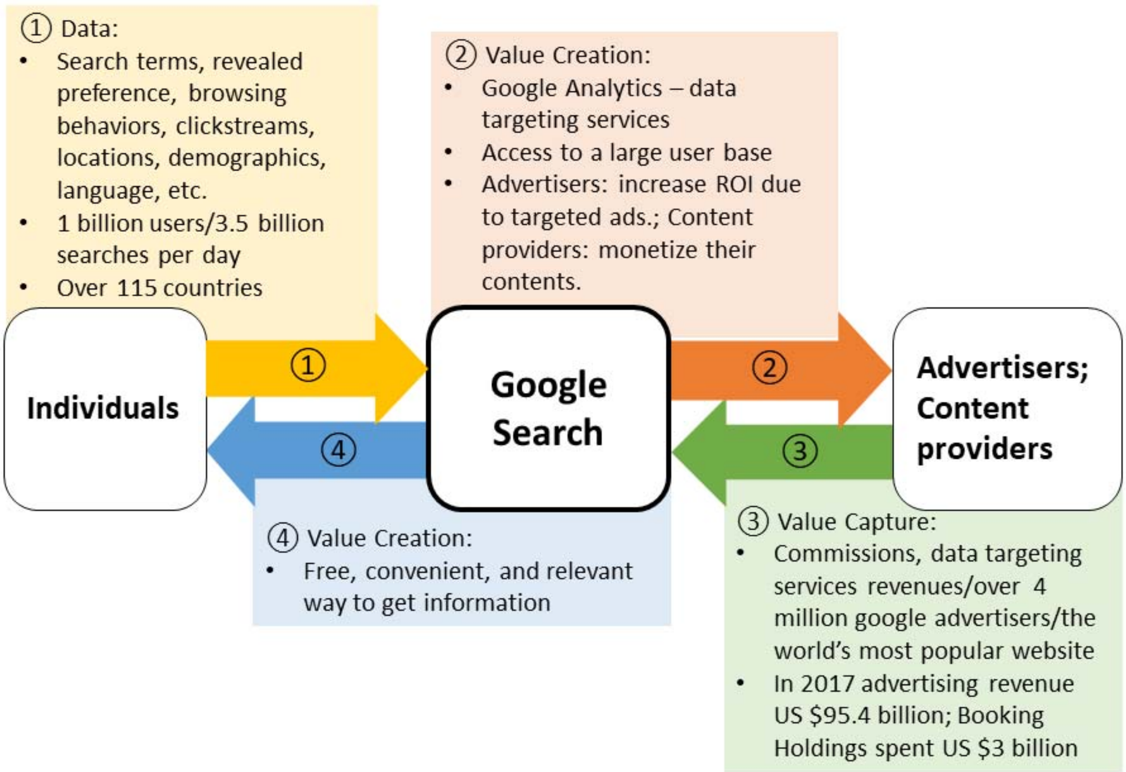

Figure 7: *Type VII: Online Search Platform*

Case Study: Google Search

In terms of data flow, Google Search collects user data on search terms, revealed preferences, browsing behaviors, locations, demographics, languages, etc. It then conducts data analytics to provide data targeting services, such as data targeted advertising and demand forecast. For example, in 2017, Booking Holdings paid Google US $3 billion for AdWords advertising. Google's revenues come mostly from data targeted advertising services, such as 87% in 2017.

3. Creation of the Value of Data, Data Value Chain, and Vertical Integration

Based on the understanding from the study of the seven major types of online platforms described in Section 2, Figure 8 summarizes how the value of data is created. Generally speaking, online platform companies collect data from users and third parties, and use two ways to monetize the data. One way is to license access to the data to clients, such as data analytics firms. Because it is highly unlikely for one company to unveil and use the full potential of data, firms without in-house data analytics capabilities and/or data monetization strategies tend to license the use of data. The other way is to provide data targeting services to clients, such as third-party sellers. This option requires internal technical skills in data fusion, data analytics, and subject matter experts to produce a data-driven business plan for a data targeting service that can produce revenue for the firm. Depending on its capabilities of data fusion and analytics and business expertise, an online platform company can offer a variety of data targeting services that produce revenues. Even within the same type of online platforms, companies can differ in the services they offer.

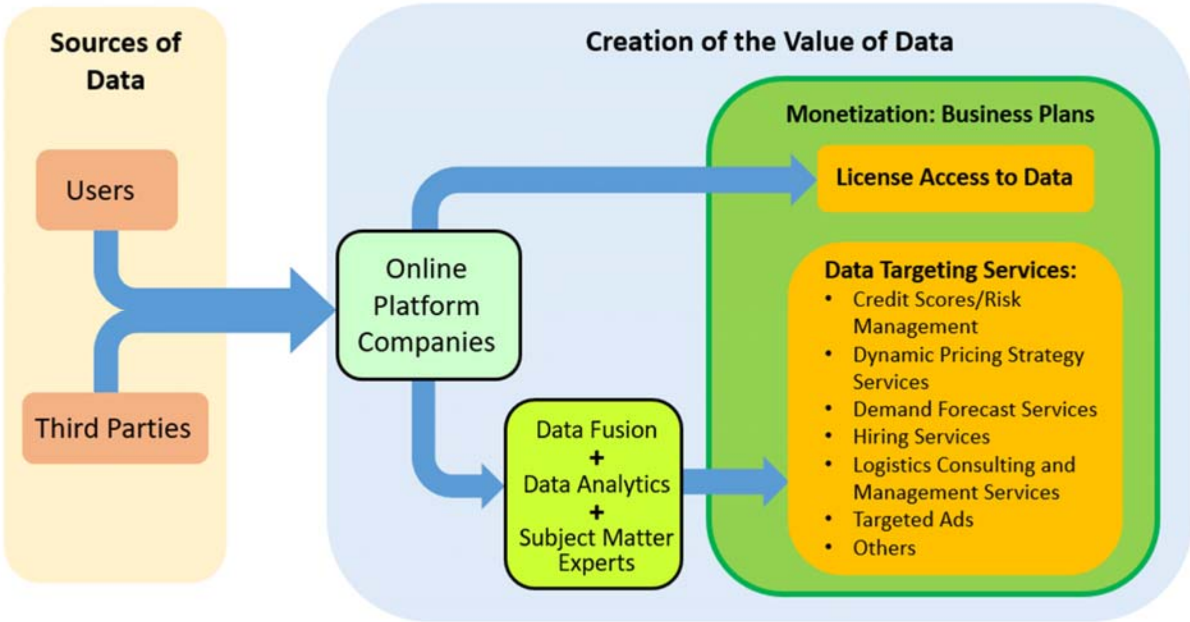

Figure 8: *Creation of the Value of Data*

Figure 9 presents the physical supply chain of data monetization and indicates that online platform companies have outsourcing options in each stage of the chain. For example, in the stage of devices, Google pays Apple traffic acquisition costs (TACs) for the rights to be the default search engine for the Safari web browser on iPhone, iPad, and Mac devices. TACs are the costs of accessing a pool of consumers and acquiring their data. While the growth of iPhone users has been slowed significantly, TACs paid by Google to Apple have increased 12-fold within 5 years and will reach US $12 billion in 2019, which is equal to over 10% of Google's data-targeted advertising revenue in 2018 (Reisinger, 2017; Williams, 2018; D'Onfro, 2018).[4] These numbers indicate the importance of the online traffic in Apple's ecosystem to Google's data-targeting services, and explain why Google purchased HTC's smartphone division in 2018 (Bergen and Sherman, 2017).

[4] Here, we use the compound quarterly growth rate of Google's advertising revenue from the first quarter of 2014 to the third quarter of 2018 to estimate its advertising revenue in the fourth quarter of 2018.

Online platform companies can also outsource data storage to cloud service providers. Cloud services, sold like utility, allows companies to tap the benefits of cost saving and high flexibility in time and capacity. Moreover, companies can even outsource the management of customer relations to firms such as Salesforce and the management of user support to firms such as ZenDesk.

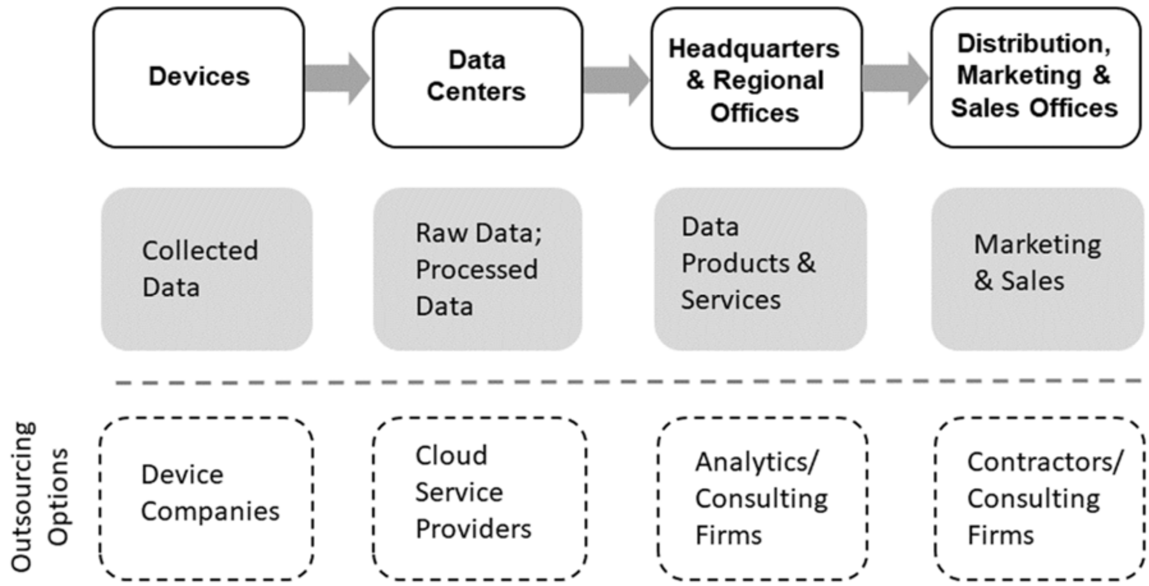

Figure 9. *Physical Supply Chain of Data Monetization*

Online platform companies perform a series of value-added activities to create value from data. Several studies have defined the big-data value chain from the engineering perspectives (e.g., Miller and Mork, 2013) but do not consider data monetization. Visconti et al. (2017) included the monetization of data as a stage that is confined only to financial planning and cash flow forecasts without organizational planning. They also consider data fusion as the main source of value creation and a new business model, a definition different from what is generally used in business management. The big-data value chains developed by these previous studies cannot well characterize the mechanisms of data monetization found in our case studies.

Based on the value creation of data and the physical supply chain identified in Figures 8 and 9, the data value chain is illustrated in Figure 10 and composed of four stages: data collection, data storage, data analytics, and data-driven business model. Data collection can include intended and unintended data collection activities. The collected data in this first stage are not limited to unstructured and uncorrelated data. Data analytics includes data processing and visualization, and the stage of data-driven business models includes business plans that guide the firm to use and monetize data. The four stages do not contribute to the creation of the value of data equally. If online platform companies cannot license the use of data and do not have capabilities in providing data-targeting services, data itself may not contain much value. Data collection and storage generate only small amounts of value and greater value is created when a firm has a data-driven business model, which is a business plan that contains monetization-driven organizational planning and cash flow forecasts. In other words, an online platform company can make some profit by licensing the use of data, but having a data-driven business model can bring the most return.

Figure 10. *Data Value Chain*

Two online platforms well exemplify the concept of data value chain and the vertical integration within the chain. Before being acquired by Google, ITA Software focused its business on the first two stages of the data value chain and licensed the use of data to companies such as

Farecast. Farecast focused its business on the last two stages by providing data-driven prediction services of airfares to consumers. After acquiring ITA Software, Google has obtained the highest degree of vertical integration in this data value chain. Another example is Twitter. Unlike Google or Amazon, Twitter lacks strong in-house data analytics capabilities and monetization strategies to vertically integrate into the data value chain fully. Even with vast amounts of data, it chose to license the use of data before 2010. After 2010, when data targeted advertising services became a popular monetization strategy, Twitter also adopted this business model to monetize its data. In the first quarter of 2018, 12.3% of Twitter's sales came from licensing the use of data, and the rest from data targeted advertising services. But, the recent growth of Twitter's earnings relies on its high-margin and fast-growing business of licensing the use of data (Bary, 2018). An understanding of the data value chain can help identify the right approach to value data.

4. Measurement of the Value of Data: Methodology and Case Studies

Since new values of data can be created through data fusion, including the fusion of different types of independent datasets, data do not depreciate differently by the type of data. New values of data can also be created through innovations in data-driven business models. These unique features of data pose challenges to valuing data. Three conventional approaches can be useful in measuring the value of data: the cost-based approach, the market-based approach, and the income-based approach (Slotin, 2018). Using the cost-based approach, such as using the salaries of data analysts and the costs of data centers, is likely to significantly underestimate the value of data. Such a problem is clear as many online platform companies outsource data centers to companies like Amazon and Microsoft to tap the benefits of cheap and flexible cloud services. To assess the market-based approach, we used a difference-in-difference method and the state space model (Varian, 2014; Scott and Varian, 2014; Brodersen et al., 2015) to assess the causality between the merger and acquisition (M&A) deals and the stock prices of the acquiring firms in our case studies. We did not find any statistically significant causality effect; however, the reason may be due to the deal sizes being too small to affect the market caps of acquiring firms. In addition, there is likely a mispriced issue because neither sellers nor buyers know the precise value of the data. Akerd and Samani (2018) point out that, during an M&A, assuming the value of data captured only by sales figures may understate the overall value of a transaction to the benefits of the buyer and to the detriment of the seller. Moreover, it is impossible to visualize all the possible ways to create values from the data in the future, especially when significant value can possibly be generated through data fusion. The usefulness of the income-based approach is rather limited, because it only applies to the cases for licensing the use of data and where transaction data are available.

Given the serious drawbacks and limitations of three conventional approaches, we consider a new approach to value data. Since most value of data is generated when a firm has a data-driven business model, this part of investments heavily rely on online platform companies' investments in business models, which can be measured by their investments in organization capital. To measure intangibles, economists generally encounter the problems that there is no arms-length market for most intangibles and that the majority of them are developed for a firm's own use. Following earlier studies, we use the sales, general, and administrative (SG&A) expense, reported in annual income statements, as a proxy for a firm's investment in organizational capital (Lev and Radhakrishnan, 2005; Eisfeldt and Papanikolaou, 2013; Brynjolfsson et al., 2018). SG&A expenditures include most of the expenditures that generate organization capital, such as employee training costs, brand enhancement costs, consulting fees, and the installation and management costs of supply chains. Because SG&A expenditures may include some items unrelated to improving a firm's organizational efficiency, Eisfeldt and Papanikolaou (2013) use five ways to validate the approach of using SG&A expenditures as a proxy for a firm's investment in organizational capital and their results show a clear support for this approach.

Moreover, the inefficiency of the investment in organization capital by definition should show in the depreciation rate of organization capital. That is, if a firm's investment in organization capital has significant inefficiency, the value of its organization capital cannot be maintained well, resulting in a higher depreciation rate of organization capital. As shown in Li (2015), across U.S. high-tech industries, market leaders in general have a smaller depreciation rate than their followers. We adopt the R&D depreciation model developed by Li and Hall (2018) to estimate the depreciation rates of organization capital for four online platform companies, including Amazon, Booking Holdings, eBay, and Google, for which public data are available. This new model is a forward-looking profit model that uses a firm's data on sales and investments in intangible capital to identify its depreciation rate of intangible capital, a new approach that can be very useful in estimating the value of intangibles.

Following Hall (1993), we use the perpetual inventory method to construct the stocks of organization capital and the associated growth rates for the four firms. The data cover the years of 2000 to 2017. Table 1 shows the estimated results (see column 4^{th}), annual commission or licensing revenue, and M&A prices associated with our case studies. For example, in 2017, Amazon's estimated annual commission derived from data is US $41.8 billion, and the estimated value of data derived from a data-driven business model is US $125 billion.

Table 1: *Measurement of the Value of Data: Case Studies*

Type of Online Platform	Company	Annual Commission or Licensing Access to Data	Value Based on Data-driven Business Model	Merger & Acquisition Price
E-commerce	Amazon	Commission Revenue: US $41.8 billion (2017) Premium Data Service Revenue: US $18 billion (2018)*	US $125 billion; Annual Growth Rate: 35% Depreciation Rate: 24.1%	
Online Sharing	Booking	US $11.8 billion (2017)	US $15.7 billion; Annual Growth Rate: 40% Depreciation Rate: 18.6%	
Fintech	Ant Financial	No public financial statement.		
Social Network Service	LinkedIn	US $2.99 billion (2015)**		US $26.4 billion by Microsoft in 2016
Matchmaking	eBay		US $16 billion; Annual Growth Rate: 30% Depreciation Rate: 26.0%	
Crowdsourcing	Waze	No public financial statement.		US $1.3 billion by Google in 2013
Search	Google		US $48.2 billion; Annual Growth Rate: 21.8% Depreciation Rate: 26.5%	

* Assume third-party sellers with annual sales over US$10 million order the premium data service. There are 19% of third-party sellers that have sales over US $10 million per year.

**Most of the revenue number from selling access to the data of its members to recruiters and sales professionals.

Tables 2 to 4 list the M&A histories of Amazon, eBay, and Google, respectively. It is clear that the purpose in most M&A cases is related to data, indicating that these online platform companies are aggressively expanding the types of the data they collect.

Table 2: *Merger & Acquisition Cases by Amazon*

Year	Acquired Firm	Purchased Price	Purchased Price /Amazon Market Cap	Business	Country	Purpose of M&A
2009	Zappos	US $1.2 billion	0.0228	Online Shoe and Apparel Retailer	USA	Data
2014	Twitch	US $0.97 billion	0.0062	Live Streaming, Streaming Video	USA	Data
2017	Whole Foods	US $13.7 billion	0.0281	Supermarket Chain	USA	Online to Offline; Data
2018	Ring	US $1.8 billion	0.0023	Video Doorbells	USA	IoT; Data
2018	PillPack	US $1 billion	0.0011	Pharmacy	USA	Data

Table 3: *Merger & Acquisition Cases by eBay*

Year	Acquired Firm	Purchased Price	Purchased Price /eBay Market Cap	Business	Country	Purpose of M&A
2002	PayPal	US $1.5 billion	0.1834	Online Payment Systems	USA	Data
2009	Skype	US $2.6 billion	0.1163	Software for Voice; Video Calls	Luxembourg	Data
2008	Bill Me Later	US $1.2 billion	0.1228	E-commerce	USA	Data
2011	GSI Commerce	US $2.4 billion	0.1423	Marketing; Fulfillment	USA	Data Analytics

Table 4: *Merger & Acquisition Cases by Google*

Year	Acquired Firm	Purchased Price	Purchased Price /Google Market Cap	Business	Country	Purpose of M&A
2006	Youtube	US $1.65 billion	0.0124	Video Sharing	USA	Data
2007	DoubleClick	US $3.1 billion	0.0212	Online Advertising	USA	Data Analytics
2012	Motorola	US $12.5 billion	0.0692	Mobile Device	USA	Data Device; Data; Reduce TACs
2013	Waze	US $1.3 billion	0.0044	GPS Navigation Software	Israel	Data
2014	Nest Labs	US $3.2 billion	0.0077	Home-security Cameras and Thermostats	USA	IoT; Data
2018	HTC Smartphone Division	US $1.1 billion	0.0015	Mobile Device	Taiwan	Data Device; Data; Reduce TACs

5. Discussion

Policy analysis on digital goods and services should consider the value of data. Lacking understanding of the value of data may hinder managers' ability to make good investment, management, and outsourcing decisions related to data. Moreover, in the rising data-driven economy, no information about the value of data may mislead investors about firm fundamentals and inhibit capital flows to innovative firms or good investment opportunities.

For example, transactions through an e-commerce platform can generate an enormous amount of data and of the value of data. Whereas a transaction itself creates a conventional economic benefit known as gains from trade, the data generated through the transaction also contains economic value. The value of such transaction data has traditionally been accumulated within a firm as firm-specific knowledge on consumers, business partners, and employees. The specific knowledge derived from the value of transaction data can then be utilized for various management departments, such as marketing, procurement, and human resource, within a firm. However, transaction data collected through online platforms are accumulated digitally and can, nowadays, easily be recombined and aggregated with other datasets. This new and unique nature of digital data allows an online platform company to utilize it to a degree that far exceeds its offline counterparts not only in scale but also in scope.

The economic value and policy implications of digital transaction data can be discussed in two scenarios, depending on whether or not the consumer's identity in a transaction is disclosed. The first scenario considers the condition where the consumer's identity is disclosed may cause potential welfare loss such as identity theft or privacy breach as reviewed in Acquisti et al. (2016).

A more subtle effect is dynamic price discrimination. For example, online platform companies can provide data targeting pricing strategy services to third-party sellers. Based on a consumer's data on transaction records and clickstreams, an online platform company can suggest a third-party seller to raise its product price for a specific customer whose data reveal that she would accept the higher price. Acquisti et al. (2016) argue that "the evidence of systematic and diffuse individual online price discrimination is, currently, scarce." However, Booking.com reportedly can on average increase the revenue of its corporate clients by 7% through its data-driven pricing strategy service. Other examples include Uber's surging pricing strategy and Amazon's dynamic pricing program.

Price discrimination can be efficient and not necessarily implies welfare loss. Price discrimination does reduce consumer surplus by reducing the margin between the consumer's willingness to pay and the purchase price. However, the reduced consumer surplus merely transfers to the firm as an increased profit, and, in a general equilibrium, the firm's increased profit is distributed to households as income. Therefore, the price-discriminated customer loses some consumer surplus, while households in the economy as a whole receive the equivalent value of additional income.

The price discrimination can also be redistributive. Provided that a high-income consumer tends to have a higher willingness to pay for goods or services, the price discrimination results in a transfer from high-income households to the representative household. If the firm's increased profits are distributed equally among households, the resulting distribution may become more egalitarian than before. However, if the increased profits accrue to only a handful of entrepreneurs, the transfer through the price discrimination does not necessarily lead to a more equal distribution. The resulting distribution depends on the ownership of emerging business models.

The second scenario considers the condition where the consumer's identity in a transaction is not disclosed or used by the service provider. For example, a consumer can enjoy the benefits gained by revealing personal attributes, but the service provider cannot identify the customer as a person. In this case, as long as the customer has an option of staying in the status quo in receiving conventional goods and services, the customer bears no surplus costs in supplying personal attributes and transaction records as an observation in data. Namely, the marginal cost of data provision is zero when anonymity is preserved.

A single data point that a data subject provides under anonymity has little value. However, a collection of them can generate a significant amount of value. That is because a collection of data can reveal statistical regularities. In this sense, an observation of customer data has a positive externality: a single observation has no value, but a collection of them potentially does. This type of externality might be called a “data network effect” mentioned by lawyers and regulators, which was dismissed by Varian (2018) as a misnomer of learning-by-doing.

The data network effect can be formulated as an externality in two-sided markets. Rochet and Tirole (2003, 2006) consider a usage externality in two-sided markets, such as the case where a video game user’s participation unintentionally benefits another user on the same platform. Similarly, a consumer’s transaction record on an online platform can benefit other consumers by improving the predictive power of the platform’s algorithm. The data-driven online platform service allows not only a consumer to search a good or service that fits personal needs more efficiently but also a third party seller to serve its target customers more effectively. That is, the transaction data accumulated through an online platform can increase the predictive power of its matching algorithm, an increased productivity in the algorithm that reduces transaction costs for both consumers and producers. In this case, the combined transaction costs needed for facilitating the same matching outcome without online platforms can be formulated as the social value of data, which in fact is an increasing function of the combined transaction costs (Appendix A). Online platform companies may capture a significant portion of the social value of data by internalizing the positive externality from the data network effect. The captured value can not only cover their investment costs in developing AI algorithms but also be very profitable. From the perspective of an online platform company, the accumulation of transaction data can increase the productivity of its matching algorithm. Moreover, data are certainly an asset, given that data are one of the key inputs for online platform companies to produce data products and services and they have earned significant revenues through monetizing data repeatedly.

In terms of the statistical value of data, Varian (2018) argues that it exhibits decreasing returns to scale, citing that an increase in the size of training data attains only diminishing gains in prediction accuracy. This is true for an objective with a single dimension. However, an extension of data to multiple dimensions may not suffer decreasing returns. For example, merging the data on two attributes of households can enhance the prediction power on both attributes. Hence, as long as a household is a "statistical" subject, a collection of data is likely to have a positive welfare effect. This is the case where a household can access the knowledge generated by the data without disclosing its identity. A person and a firm can share the increased value added.

In addition to the business-to-consumer redistribution effect we have discussed above, a business-to-business redistribution effect may arise as a result of an accumulation of valuable data. An accumulation of data and an introduction of new online platform business models may cause business-stealing effects and hurt incumbents, a creative destruction phenomenon analyzed by Li, Nirei, and Yamana (2018) for the U.S. and Japan's hospitality and transportation industries. In our analysis, an introduction of new online platform business models speeds up the obsolescence of conventional business models. As a result, the accumulated intangible capital of conventional businesses depreciates faster.

The creative destruction process has a redistributive effect but not necessarily implies welfare loss. For example, an online platform company can accumulate an enormous amount of data to gain a great competitive advantage and render the business models of its conventional counterparts obsolete. Consequently, the consumer surplus, income, and rent generated by conventional businesses decline, but those generated by new businesses increase. Li, Nirei, and Yamana (2018) estimate a lower bound of the redistributed value by conducting case studies on the gains and losses of firms' market valuations.

Lastly, data are information goods and act like knowledge. Therefore, many arguments on knowledge can also apply to data. A negative externality on data production is a duplicated investment, or the “stepping on toes effect” (Jones and Williams, 2000). A firm’s data generation may overlap with those by other firms, which causes a duplicated investment and pure welfare loss. In fact, because data is non-rival, there are potentially large gains by sharing data, the information externality effect of data (Jones and Tonetti, 2018). However, the rising data privacy and security concerns have become the main obstacle for firms to share data. To cope with the rising concerns, firms have invested in new technologies allowing individuals and organizations across industries to share data to gain greater insights from larger datasets and co-develop new data products and services without sacrificing privacy. For example, JPMorgan invests in Inpher, whose zero-knowledge computing technology allows analysis on encrypted data (Castellanos, 2018). Moreover, as data are emerging as the key differentiator for companies in the AI race, data have become a concern of impeding the entry of SMEs. New market developments may mitigate the concern. Data brokers like Experian sell personal data to individuals and firms at competitive prices, and firms like Nasdaq make efforts to establish markets for data (Wigglesworth, 2018; Ram and Murgia, 2019). Furthermore, the world’s first data exchange market, for trading data such as court and medical records, started operation in China in 2015. Members with government approvals to trade, including Alibaba and Tencent, increased 20-fold to more than 2000 within three years (Kang, 2015; China Daily, 2018). More importantly, data sharing across countries matters because there is a limitation of a country’s data on AI development. For example, the patterns recognized in Chinese consumer data will not necessarily be able to apply to consumers in the U.S. (Dvorak, 2018).

6. Conclusion

Online platform companies are data companies, normally physical-asset-light but highly profitable. They exchange "free" digital goods and services for consumer data. Data are an intangible, and its value is very difficult to measure. In this study, we propose a way to estimate the value of data for several representative firms in the seven types of online platforms. Our initial results indicate that data can have enormous value: for example, in 2017, the value of Amazon's data can account for 16% of Amazon's market valuation and has an annual growth rate of 35%.

Online platforms can differ in the underlying business model, which determines what types of data they collect, how data flow within online platform networks, how online platform companies monetize data, and what consumers gain from exchanging their data. We select seven major types of online platforms, and conduct case studies to understand the data activities related to them and to examine whether the data monetization strategies vary by the type of online platform. Based on the understanding from those case studies, we derive a flow chart to show the steps by which online platform companies create the value of data. We also present the data value chain to demonstrate the value-added activities involved in each step.

We find that online platform companies can differ in the degrees of vertical integration in the data value chain, a difference which determines how they monetize data and how much economic benefits they can capture. Online platform companies with in-house data analytics capabilities and monetization strategies can produce much greater values of data than do those that outsource data analytics work. More importantly, online platform companies are at the forefront of AI adoption, and data are emerging as a key differentiator in the AI race. Greater vertically integrated online platform companies can benefit more from data. Their businesses can be strengthened by the virtuous cycle between AI's relationship with data (Lee, 2018). That is, more data can lead to better digital goods and services, which in turn attracts more users to their online platforms, generating even more data that further improve their digital goods and services.

Currently, there is no definitive answer to the welfare implications of online platforms and data. For example, on the one hand, we find that online platform companies can offer data-driven pricing discrimination strategies for its corporate clients to maximize their revenues; on the other hand, the households in the economy as a whole can receive the equivalent value of additional income. The price discrimination can also be redistributive, and the resulting distribution depends on the ownership of emerging business models.

Moreover, there is a positive externality from the data network effect derived from consumer data. Because a collection of data can reveal statistical regularities, a consumer's transaction record can benefit other consumers by improving the predictive power of the platform's matching algorithm, an increased productivity in the algorithm that reduces transaction costs for both consumers and producers. In this case, the combined transaction cost needed for facilitating the same matching outcome without online platforms is the social value of data. Online platform companies may capture a significant portion of the social value of data by internalizing the positive externality from the data network effect. The captured value can not only cover their investment costs in developing AI algorithms but also be very profitable.

Nevertheless, an accumulation of data and new online platform business models may cause business-stealing effects and speed up the obsolescence of conventional business models. Data have also become a concern of impeding the entry of SMEs and companies may make duplicate investments in data. Since there are potentially large gains by sharing data, new technologies facilitating data sharing without sacrificing privacy and market institutions such as data exchange market can help mitigate the concern and encourage innovation.

Online platforms are evolving rapidly. To seek growth, online platform companies may expand their business models to cover multiple types of online platforms. The degree of hybrid online platforms can vary across countries. Compared with less hybrid U.S. counterparts, some Chinese online platform companies have developed super online platforms that bundle many online platform functionalities similar to Facebook, Uber, Expedia, PayPal, Amazon and more combined, an outcome called "The App Constellation Model" (Lee, 2018). Moreover, though online platforms in general have two ways to monetize data, they differ in the type of collected data and the difference may affect their diversification strategies. For example, compared with LinkedIn, Amazon can more easily enter the data-targeted advertising market because of its greater advantage in real-time browsing and shopping data. More research is needed to understand the impacts of the rapidly evolving trend of online platforms on areas including data collection, market competition, and consumer welfare.

Lastly, data is new oil. At present, the pipelines of new oil are controlled by online platform companies. In the future, blockchain technology may allow consumers to have their own pipelines, to take control of their ownership, and to decide whether and how to sell personal data to companies. However, since the value of data is created by companies and depends on data analytics and the associated business model, consumers lack the knowledge to value their personal data. Nevertheless, how fast various industries can adopt blockchain technology may affect the future competition of online platform companies. Currently, data volume doubles every three years (Mayer-Schönberger and Cukier, 2014), but trends such as the fifth generation of mobile networks and the Internet of Things are rapidly accelerating the accumulation speed of data types and volume. Therefore, the capability to value data is very important not only at the firm level but also at the national level. At the firm level, a proper valuation of data is important for firms to derive important investment and outsourcing decisions on data, decide how to monetize data, and gain a competitive edge through data. At the national level, it is important for National Accounts to incorporate this increasingly important new asset into the calculation of GDP and productivity growth. Moreover, countries differ in the ownership and collection right of personal data such as Europe's strict General Data Protection Regulation and China's extreme openness of personal data. Additionally, the U.S. allows foreign firms to collect personal data domestically but China prohibits it. How do the differences in data policy affect trade? Given the virtuous cycle between AI's relationship with data, the degree of openness of a country's data policy may affect relative competitiveness between domestic and foreign firms. Therefore, the valuation of data will provide important policy implications for innovation, investment, trade, and growth.

References

Acquisti, A., Taylor, C., & Wagman, L. 2016. The Economics of Privacy, *Journal of Economic Literature*, 54(2), June, pp: 442-92.

Akred, J., & Samani, A. 2018. Your Data Is Worth More Than You Think. *MIT Sloan Management Review*. January 18th.

Agrawal, A., Gans, J., & Goldfarb, A. 2018. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Review Press.

Bary, E. 2018. Twitter Earnings Growth Relied on Selling User Data, *Market Watch*, April 25th.

Beck, M., & Libert, B. 2019. The Machine Learning Race Is Really a Data Race, *MIT Sloan Management Review*, Winter.

Bergen, M. and Sherman, A. 2017. Google Is Close to Buying HTC Assets to Bolster Hardware, *Bloomberg*, September 20th.

Bond, S. 2018. Amazon's Ever-increasing Power Unnerves Vendors, *Financial Times,* September 21st.

Brodersen, K.H., Gallusser, F., Koehler, J., Remy, N., and Scott, S.L. 2015. Inferring Causal Impact Using Bayesian Structural Time-series Models. *Annals of Applied Statistics*, 9, pp: 247-274.

Brynjolfsson, E., Collis, A., Diewert, W.E., Eggers, F., & Fox, K.J. 2018. The Digital Economy, GDP and Consumer Welfare: Theory and Evidence, *the Sixth IMF Statistical Forum proceeding papers*, November, Washington, DC.

Brynjolfsson, E., Rock, D., & Syverson, C. 2018. The Productivity J-Curve: How Intangibles Complement General Purpose Technologies, *National Bureau of Economic Research Working Paper No. 25148*, October.

Castellanos, S. 2018. JPMorgan Invests in Startup Tech that Analyze Encrypted Data, *Wall Street Journal*, November 13th.

Chen, Y., Dai, T., Korpeoglu, C., Korpeoglu, E., Sahin, O., Tang, C., & Xiao, S. 2018. Innovative Online Platforms: Research Opportunities, *Manufacturing & Service Operations Management*, forthcoming.

China Daily. 2018. After More Than Three-year Exploration, Guiyang Big Data Exchange Market Innovates New Data Exchange Model, *China Daily*, July 4th, China.

Cohan, P. 2013. Google to Spite Facebook, Buy Wave For $1.3 Billion, *Forbes*, June 9th.

Demunter, C. 2018. Towards a Taxonomy of Platforms in the Collaborative Economy: Outcomes of a Workshop on Measuring the Collaborative Economy, presented at the 2018 OECD Workshop on Online Platforms, Cloud Computing, and Related Products, September 6th, OECD, Paris.

D'Onfro, J. Alphabet Stock Sinks on Revenue Miss, *CNBC*, October 25th.

Dvorak, P. 2018. Which Country Is Winning the AI Race – the U.S. or China? *Wall Street Journal*, November 12th.

Eisfeldt, A., & Papanikolaou, D. 2013. Organizational Capital and the Cross-Section of Expected Returns, *Journal of Finance*, 4, August, pp: 1365-1406.

Frier, S. 2018. Is Apple Really Your Privacy Hero? *Bloomberg Businessweek*, August, 8th.

Hall, B.H. 1993. The Stock Market's Valuation of R&D Investment During the 1980's, *American Economic Review*, 83(2), Papers and Proceedings of the Hundred and Fifth Annual Meeting of the American Economic Association, May, pp: 259-264.

Hartmans, A. 2017. Airbnb Has More Listings Worldwide than the Top Five Hotel Brands Combined, *Business Insider*, August 10th.

Hathaway, I., & Muro, M. 2017. Ridesharing Hits Hyper-growth, Brookings.

Johnson, M.W. 2018. *Reinvent Your Business Model: How to Seize the White Space for Transformative Growth*. Harvard Business Review Press.

Jones, C.I., & Williams, J.C. 2000. Too Much of a Good Thing? The Economics of Investment in R&D, *Journal of Economic Growth*, March, 5(1): 65-85.

Jones, I.J., & Tonetti, C. 2018. Nonrivalry and the Economics of Data, *Stanford GSB working paper*.

Kang, Y.P. 2015. The World's First Data Exchange Market with Annual Transaction Amount up to US $29.4 Million, *Business Weekly*, 1438, June, Taiwan.

Lee, K.F. 2018. *AI Superpowers: China, Silicon Valley, and the New World Order*. Houghton Mifflin Harcourt. September 25th.

Lev, B., & Radhakrishnan, S. 2005. The Valuation of Organizational Capital. In *Measuring Capital in a New Economy*, Corrado, C., Haltiwanger, J., Sichel, D. (eds). Chicago: National Bureau of Economic Research and University of Chicago Press: Chicago, IL: 73-99.

Li, W.C.Y. 2015. Organizational Capital, R&D Assets, and Offshore Outsourcing, Working Paper, U.S. Bureau of Economic Analysis.

Li, W.C.Y., & Hall, B. H. 2018. Depreciation of Business R&D Capital, *Review of Income and Wealth,* https://doi.org/10.1111/roiw.12380.

Li, W.C.Y., Nirei, M., & Yamana, K. 2017. Sharing Economy in Japan and the United States, *2017 SEM proceeding papers,* July, Cambridge, the United States.

Li, W.C.Y., Nirei, M., & Yamana, K. 2018. Creative Destruction in Organizational Capital: Evidence from the Online Platform Economy in Japan and the United States, *the 2018 IARIW proceeding papers,* August, Copenhagen, Denmark.

Mayer-Schönberger, V., & Cukier, K. 2014. *Big Data*. Mariner Books: Boston and New York.

Miller, H.G., & Mork, P. 2013. From Data to Decisions: A Value Chain for Big Data. *IT Professional*.

Mims, C. 2018. How Amazon's Ad Business Could Threaten Google and Facebook, *Wall Street Journal*, January 28th, 2018.

Molla, R. 2017. Amazon Could Be Responsible for Nearly Half of U.S. E-commerce Sales in 2017, *Recode*, October 24th.

OECD. 2018a. *Online Platforms: A Practical Approach to Their Economic and Social Impacts*. Paris.

OECD. 2018b. *Online Platforms: Annex Document*. Paris.

Ovenden, J. 2016. Customer Analytics Case Study: eBay, *Innovation Enterprise Channels*, June 6th.

Ram, A., & Murgia, M. 2019. Data Brokers: Regulators Try to Rein in the 'Privacy Deathstars,' *Financial Times*, January 7th.

Reisinger, D. 2017. Here's How Many iPhone Are Currently Being Used Worldwide, *Fortune*, March 6th.

Rochet, J.C., & Tirole, J. 2003. Platform Competition in Two-sided Markets. *Journal of the European Economic Association*, Volume 1(4):990-1029.

Rochet, J.C., & Tirole, J. 2006. Two-sided Markets: A Progress Report. *The RAND Journal of Economics*, Volume 37(3):645-667.

Scott, S. L., & Varian, H. R. 2014a. *Bayesian Variable Selection for Nowcasting Economic Time Series*. Economic Analysis of the Digital Economy.

Scott, S. L., & Varian, H. R. 2014b. Predicting the Present with Bayesian Structural Time Series. *International Journal of Mathematical Modelling and Numerical Optimisation*.

SEC. 2012. Facebook, Inc. 10-K Annual Report for the Fiscal Year Ended December 31, 2012.

SEC. 2017. Booking Holding Inc. (BKNG) SEC Filing 10-K Annual Report for the Fiscal Year Ended December 31, 2017.

Slotin, J. 2018. What Do We Know about the Value of Data? May 3rd.

Statista. 2018. Percentage of Paid Units Sold by Third-party Sellers on Amazon Platform as of 2nd Quarter 2018.

Tefis Team. 2017. Facebook's Strong Ad Revenue Growth to Continue, *Forbes*, October 30th.

Varian, H.R. 2014. Big Data: New Tricks for Econometrics, *Journal of Economic Perspectives*, 28(2), Spring, pp: 3-28.

Varian, H. 2018. Artificial Intelligence, Economics, and Industrial Organization, *NBER Working Paper,* No. 24839.

van de Ven, P. 2018. Online Platforms in a Digital Economy Satellite Account, presented at the 2018 OECD Workshop on Online Platforms, Cloud Computing, and Related Products, September 6th, OECD, Paris.

Visconti, R.M., Larocca, A., and Marconi, M. 2017. Big data-Driven Value Chains and Digital Platforms: from Value Co-Creation to Monetization, working paper.

Whitmore, G. 2018. How Google Travel Is Changing the Game with Google Flights, Google Trips, and More, *Forbes*, February 20th.

Wigglesworth, R. 2018. Nasdaq Moves into 'Alternative Data' with Quandl Acquistion, *Financial Times*, December 4th.

Williams, R. 2018. Goldman: Apple Charge Google $12B to Be Default Search Engine in 2019, *Mobile Marketer*, October 1st.

Yin, Y.H. 2018. The Most Successful Internet Company after Amazon – From Netherlands, *Business Weekly*, 1600, July, Taiwan.

Appendix A: Social Value of Data and Transaction Cost

Below we present a simple model to illustrate the relationship between the social value of data and transaction costs without online platforms.

Suppose that there are a continuum of consumers $i \in [0,1]$ and a continuum of producers $j \in [0,1]$. When i and j are matched, a joint surplus value $a(i,j)$ is generated. Suppose that the surplus value function has a structure: $a(i,j) = (1 + i - j) mod\, 1$. Thus, $\sup a(i,j) = 1$ is achieved at $j \searrow i$.

First, we consider the situation where consumers cannot identify the types of producers. A consumer is randomly matched with a producer. The consumer can accept the match and leave the matching procedure. If the consumer rejects the match, the consumer loses transaction cost $c < 1/2$ and draws another match. The consumer maximizes the expected surplus value, which satisfies the recursive relation:

$$v = \max_x \int_x^1 p(\,a) a da + \Pr(a < x)(v - c)$$

where x is the consumer's reservation surplus value. If a is above x, the consumer accepts the match; otherwise, rejects. For the maximization problem, the optimal choice of x satisfies the first order condition:

$$-p(x)x + p(x)(v - c) = 0,$$

namely $x = v - c$. Note that the producer is distributed uniformly over [0,1]. By plugging the value of x back to the recursive equation, we obtain $x(1 - x) + c = (1 - x^2)/2$. By solving this equation, we obtain $x = 1 - \sqrt{2c}$ and $v = 1 + c - \sqrt{2c}$. If the transaction cost c drops to 0, the value of v increases to 1. In this scenario, the consumer will draw matches indefinitely until getting the best match.

Next, we consider the situation where consumers can identify the types of producers and hence a joint surplus value, a, arbitrarily close to supremum 1 can be achieved. The resulting increase in the expected joint value when the types of producers are known is $\sqrt{2c} - c > 0$. The increase in the expected joint value arises from (1) the reduction in the expected transaction costs and (2) the increase in reservation value ($x < 1$) which leads to increased expected surplus before netting the transaction costs.

Now, when an online platform collects transaction records from both consumers and producers, the collected data can reveal statistical regularities which can then reveal the joint surplus value structure $a(i, j)$. Because statistical regularities revealed from collected data can increase the predictive power of the online platform's matching algorithm, consumers may achieve a supremum match without incurring costs. Therefore, even though there is no cost for a consumer to provide such a transaction record, the collection of records has a social value $\sqrt{2c} - c$, the increased expected joint surplus value when online platforms exist. The social value of data is an increasing function of the transaction cost c without an online platform. This social value of data arises from the positive externality of consumer data: a single observation has no value but a collection of them does.