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)[1,2]

Makoto Nirei (University of Tokyo)

Kazufumi Yamana (Kanagawa University)

Date: November 8, 2018[3]

Abstract

The Facebook-Cambridge Analytica data scandal shows evidence that there is no such thing as a free lunch in the digital world. Online platform companies exchange "free" goods and services for consumer data, reaping potentially significant economic benefits by monetizing those data. Phrases such as "free goods" are misnomers. Welfare analysis on digital goods or services without considering the value of data can mislead policy analysis. In this research, we classify online platforms into eight major types based on underlying business models, and conduct case studies to analyze data activities related to each type. We show how online platform companies take steps to create the value of data, and present the data value chain to show how the value of data varies by 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 benefits they can capture. Unlike R&D that may depreciate due to obsolescence, data can produce new values through data fusion, a unique feature that can create unprecedented challenges in measurements. Our initial estimation shows that the value of data can be tremendous. Moreover, online platform companies can capture most benefits of the data, because they create the value of data and consumers lack knowledge to value their own data. Lastly, the Internet of Things, the trend of 5G, and the emerging online-to-offline transition are accelerating the speed of data accumulation. The valuation of data will have important policy implications for investment, trade, and growth.

[1] The views expressed are those of the author and do not necessarily reflect those of the U.S. Department of Commerce or the Bureau of Economic Analysis.

[3] We thank Sadao Nagaoka, Dylan Rassier, Paul Roberts, Rahul Telang, and participants in the *2018 IP Statistics for Decision Makers Conference* 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 different types of online platforms. For example, online resource sharing platforms, such as Uber and Airbnb, increase the efficiency of underutilized assets and lower the consumption prices of the services. Online e-commerce platforms, such as Amazon Marketplace, have greatly reduced transaction costs for many small and medium sized enterprises to sell products across states and across borders. Online platforms, mostly created and run by young companies, are physical-asset-light but have grown fast and deeply disrupted many industries. A prominent example is Airbnb, a company that has only 1.7% of the employee size of Marriott International, the world's largest hotel chain, but more listed properties than the top five global hotel brands combined (Hartmans, 2017). Moreover, the size and scale of online platforms have been growing rapidly. For example, a recent Brookings' study based on Census data shows that the U.S. ridesharing service has been experiencing a hyper-growth rate and is predicted to take over the taxi services in the near future (Hathaway and Muro, 2017). The trend exhibits a fast growth of the online platform economy across the globe. According to the European Commission (2015), between 2001 and 2011, online platforms accounted for 55% of GDP growth in the U.S. and 30% of GDP growth in the European Union.

Most online platforms have been providing digital goods and services to consumers at seemingly zero monetary cost, and economists have been trying to measure the welfare effects related to "free" digital goods and services. For example, by conducting experiments on how much monetary compensation a respondent will accept to give up a range of free digital goods and services for a certain amount of time, Brynjolfsson et al. (2018) estimate that Wikipedia in the U.S.

alone creates US $50 billion consumer surplus per year. However, the Facebook-Cambridge Analytica data scandal shows evidence that there is no such thing as a free lunch in the digital world (Bloomberg, 2018). Consumers, in fact, exchange their personal data ownership for "free" digital goods and services. As large data holders, online platform companies like Google and Facebook can reap potentially significant economic benefits by conducting analytics on their data 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. These companies have collected copious amounts of rich data through their online platforms, monetized the data, and created a tremendous amount 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 an average 95% in the past three years (SEC, 2017). At its headquarters in Amsterdam, the Netherlands, Booking has 1,800 engineers, which accounts for 90% of its employees (Yin, 2018). While being a data company, Booking outsources its data centers to take advantage of cheap cloud computing services (SEC, 2017). Another example is Facebook: when it went public in 2011, the value of its total assets was reported at US $6.3 billion, but the market valuation reached as high as US $104 billion (SEC, 2012). The huge gap between the value of its total assets and the market valuation, US $ 97.7 billion, should be accounted for by the value of its intangible assets, including the value of data. Facebook provides free social media services to users and in exchange collects data from users. It conducts analytics on user data to provide third parties with data targeting services, currently mainly data targeted advertising. In 2017, Facebook reported an advertising revenue of US $39.9 billion, and their 40% of the annual revenue growth rate was mainly driven by growth from advertising revenue (Forbes, 2017).

How big is the value of data possessed by online platform companies? Two examples can help us visualize the size of the value of data. The first example is Apple. Recently, *Bloomberg Businessweek* reports that because app developers need to pay Apple 30% commission of their sales to get access to Apple's consumer data, Apple has earned US $42.8 billion in revenue in the past decade by sharing the consumer data with app developers (Frier, 2018). The second example is ITA Software versus Farecast. ITA Software is a large airline reservation network and collects 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 the trend of airfares (Mayer-Schonberger 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 difference in the acquisition prices of the two firms indicates that data can potentially be more valuable than analytics capabilities. As explained by Lee (2018), in the age of AI implementation, data will be the core to govern the overall power and accuracy of an algorithm once computing power and engineering talent have reached a certain level. Moreover, how firms utilize their data analytics to monetize the data relies on the underlying business models, as shown in Google's purchase. When Google purchased ITA Software, it should have a business plan on how to monetize the data. In 2011, three years after the purchase of ITA Software, Google launched the Google Flights online service, which has become the most popular flight search online platform in the United States (Whitmore, 2018).

The substantial market valuation of data shown in the ITA versus Farecast example highlights the importance of measuring data activities related to online platforms. The measurement of the market valuation of data at the national, industry, and corporate levels will be important for economic statistics and public policy, such as trade or growth policy for data-driven digital goods and services, as well as corporate strategies, such as whether to outsource the work of data analytics.

Online platforms can differ in their underlying business models. The underlying business models determine what type of data they collect, how data flow within online platform networks, how online platform companies monetize the data, and what consumers can get by exchanging their data. Therefore, we need to classify online platforms based on their underlying business model so that we can explore the data activities related to them and measure the value of data online platform companies possess. Moreover, unlike tangible capital, data, one type of intangibles, do not wear and tear. Furthermore, unlike regular intangibles, such as R&D assets, which may depreciate due to obsolescence, the aggregation and recombination of data can create new value. These unique features of data pose measurement challenges to capture the value of data. Lastly, it is well known that it is difficult to get data and/or information from online platform companies (Demunter, 2018).

In this paper, we classify online platforms into eight major types based on underlying business models, and conduct case studies to analyze data activities associated with each type of online platforms. Our analysis considers the dimensions of business models, data flow, value creation for consumers, value creation for third parties, and how online platform companies monetize their data. We show how online platform companies take steps to create the value of data, and present a data value chain to show how the value of data varies by 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 benefits they can capture. Unlike R&D that may depreciate due to obsolescence, data can produce new value through data fusion, a unique feature that can create unprecedented challenges in measurement. Our initial estimation shows that the value of data can be tremendous. Moreover, online platform companies can capture most benefits of the data, because they create the value of data and consumers lack knowledge to value their own data.

In what follows, section 2 describes the typology of online platforms and data activities associated with online platforms. Section 3 discusses the mechanism that creates the value of data, data value chain, and measurement issues related to the value of data. Section 4 proposes approaches to estimate the value of data related to the case studies in the eight types of online platforms. Section 5 discusses welfare. Section 6 concludes.

2. Typology of Online Platforms and Data Activities

2.1 Definition of Online Platform

In this research, we adopt the European Commission (2015) broader definition of online platform. The European Commission (2015) defines an online platform as "an undertaking operating in two- or multi-sided markets, which uses the internet to enable interactions between two or more distinct but interdependent groups of users so as to generate value for at least one of the groups. Certain platforms also qualify as intermediary service providers." Note that because its 2015 definition is too general for regulatory purposes, the European Commission (2017) refines its original definition by specifying the underlying business models. Online platforms are defined as digital platforms that "enable consumers to find online information and businesses to exploit the advantages of e-commerce. Online platforms share key characteristics, including 1) the use of information and communication technologies to facilitate interactions between users, 2) the collection and use of data about these interactions, and (3) network effects which make the use of the platforms with most users most valuable to other users."

2.2 Typology of Online Platforms and Data Activities

A few studies have attempted to provide a typology of online platforms (Demunter, 2018; van de Ven, 2018; Chen et al., 2018). In this research, our goal is to classify online platforms in order to help measure the value of data related to them. Based on the underlying business models, we classify online platforms into eight major types and conduct case studies to examine their data activities. Each type of online platform can be further classified into sub-categories in terms of the type of demand and supply sides, i.e., business to business (B2B), business to consumer (B2C), or consumer to consumer (C2C). Due to the serious limitation in publicly accessible information, we focus on the companies for which some public data or reports are available. In addition, to seek growth, online platform companies may expand their existing business models from covering one type of online platform to multiple types. This is similar to the situation where a firm conducts businesses in multiple industries. That is, we may see online platform companies develop hybrid online platforms that cover several basic types of online platforms that we classify below.

Despite some complications that can go in the classification of online platforms, it is useful to establish a set of basic types of online platforms, aiding the understanding underlying business models and the data activities involved. We identify eight major types of online platforms as follows:

Type I: E-commerce Online Platform
Type II: Online Resource Sharing Platform
Type III: E-financial Service Online Platform
Type IV: Online Social Network Service Platform
Type V: Online Auction or Matching Platform
Type VI: Online Competitive Crowdsourcing Platform

Type VII: Online Noncompetitive Crowdsourcing Platform

Type VIII: Online Search Platform

For each type of online platforms, we conduct a case study to examine its underlying business model, data flow, value creation for consumers, value creation for third parties, and how an online platform company monetize its data.

2.2.1 Type I: E-commerce Online Platform

The first type of online platforms is the e-commerce online platform, and our case study is Amazon Marketplace (see **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 manner.

Amazon charges third-party sellers a commission of approximately 30% of their sales (WSJ, 2018). The commission pays for not only the cost of accessing one of the world's large e-commerce markets but also the cost of "basic" access to Amazon's consumer data. For example, when a consumer purchases a goods by cash in an offline supermarket, the supermarket and the third-party seller that offers the goods 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 goods online through Amazon Marketplace, not only Amazon but also the third-party seller can acquire the consumer data. Nonetheless, there is a difference in terms of the degree and the details of the data. The third-party seller can get the data displayed in the transaction; on the other hand, 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.

In terms of data flow, Amazon collects data on clickstreams, purchases, reviews, and locations from consumers.[4] Then, it conducts analytics on those data 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 logistics consulting services such as where to build the warehouse. Bond (2018) reports that Amazon offers corporate clients a premium data services, which include demand and trend forecasts, and the price for such premium data service starts from $100,000 per year. In addition, Amazon gathers information of consumer price sensitivity by funding discounts on third-party products (Bond, 2018). 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 market share of the U.S. e-commerce market (Molla, 2017). In addition, 50.5% of its e-commerce sales are conducted through its third-party sellers on Amazon Marketplace (Statista, 2018). Given the fact that the 2017 sales for Amazon Marketplace is US $139.5 billion and that Amazon charges third party sellers a 30% commission on their sales, Amazon's annual revenue from the commission is estimated around US $41.8 billion (Amazon 10K report). While growing fast, Amazon's annual data targeted advertising revenue amounted only to US $3 billion in 2017, or a mere 2.2% of its total revenue in that year. Compared to Facebook and Google, Amazon does not rely on advertising revenue.

[4] 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.

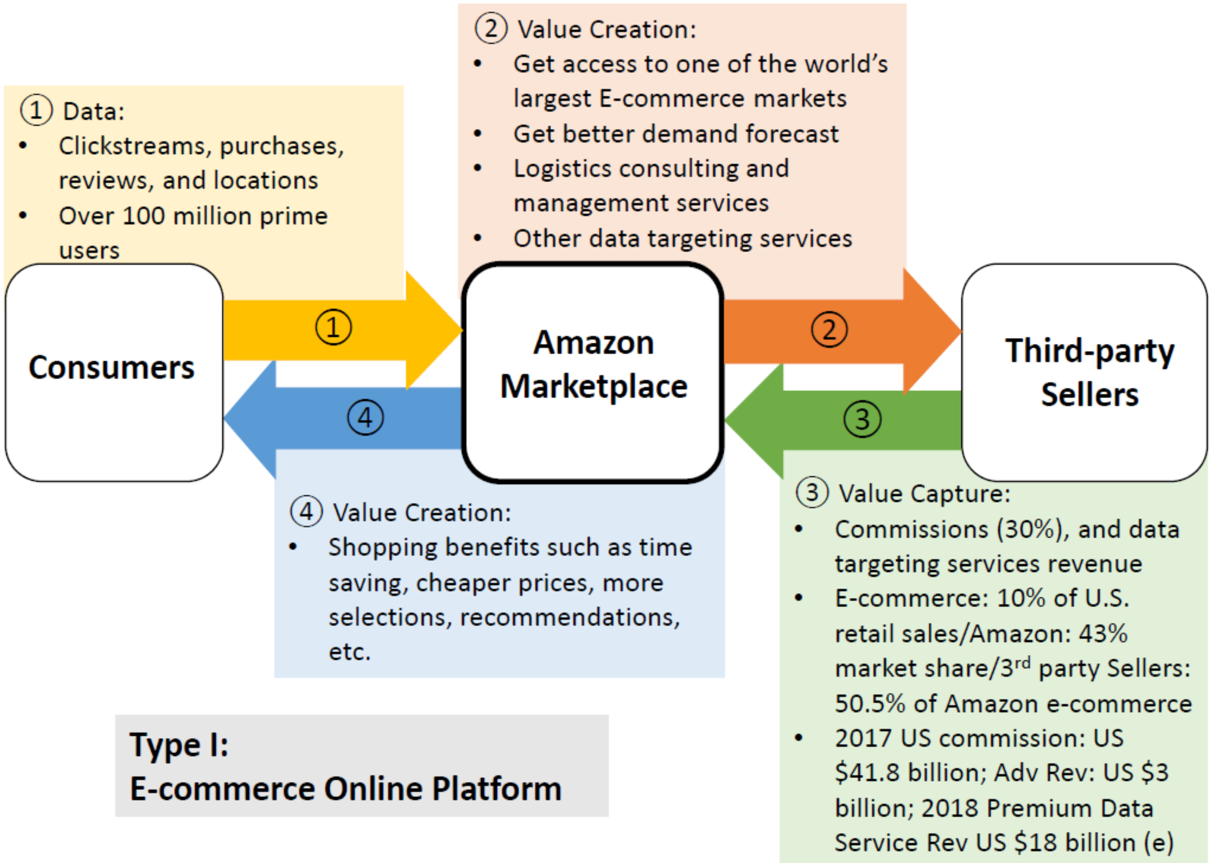


Figure 1: *Type I: E-commerce Online Platform*

Case Study: Amazon Marketplace

2.2.2 Type II: Online Resource Sharing Platform

Type II is the online resource sharing platform, and our case study is Booking.com (see **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. It 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 analytics on these data to provide third-party sellers 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). In 2017, Booking.com had 28.9 million listed properties at 1,137,791 destinations in 229 countries. The total number of its listed available private rooms is larger than that of Airbnb. Given the fact that Booking.com charges third-party sellers a 15% commission on their sales revenue, the 2017 revenue from commissions alone is around US $11.8 billion. Booking Holdings, as an online platform company, is a data company. At its Amsterdam headquarters, 90% of employees are software engineers. The company outsources its data centers and benefits from cheap cloud computing services, another business strategy that makes it physical-asset-light but extremely profitable. Based on public financial statements, its gross profit margin was 98% in 2017 and maintained an average of 95% in past three years.

What is the difference between Booking.com and Marriott International, the world's largest hotel chain and also 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, the number of its listed properties is far fewer than those of online travel platform companies like Airbnb and Booking.com, which can reach a much broader range of property owners. Moreover, its 2017 gross profit margin, 16%, is far less than the 98% of Booking Holdings.

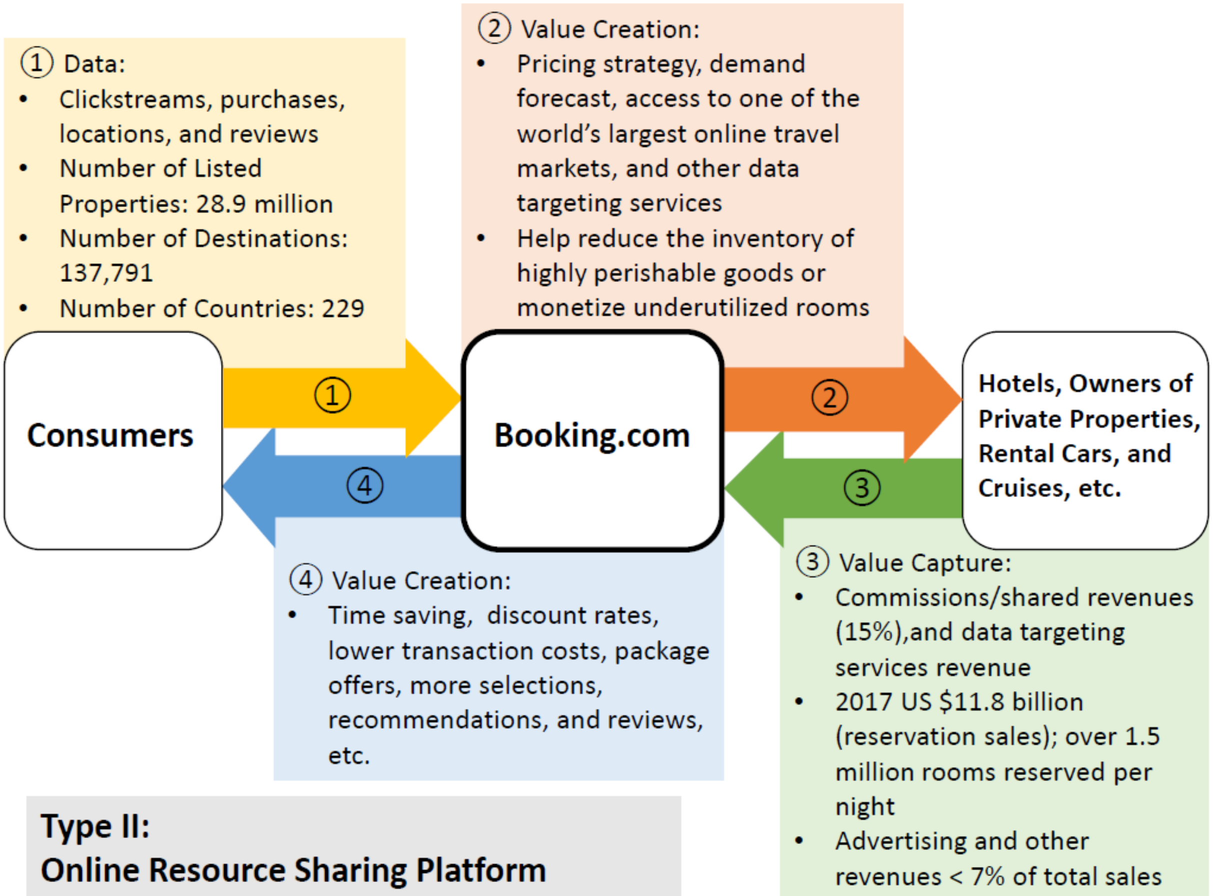


Figure 2: *Type II: Online Resource Sharing Platform*

Case Study: Booking.com

2.2.3 Type III: E-financial Service Online Platform

Type III is the e-financial service online platform, and our case study is Ant Financial (see **Figure 3**). Ant Financial is China's biggest 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 get access the credit where 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. It conducts analytics on those data to provide services to financial institutions such as credit ranking. Currently, its third-party institutions include more than 200 banks, 60 insurance companies, and over 700,000 stores. In addition, it offers data targeting demand and credit scoring services to vendors such as hotels. The reported revenue from Alipay, its online payment platform, is US $1 billion. It should be noted that consumers pay zero cost to Alipay.

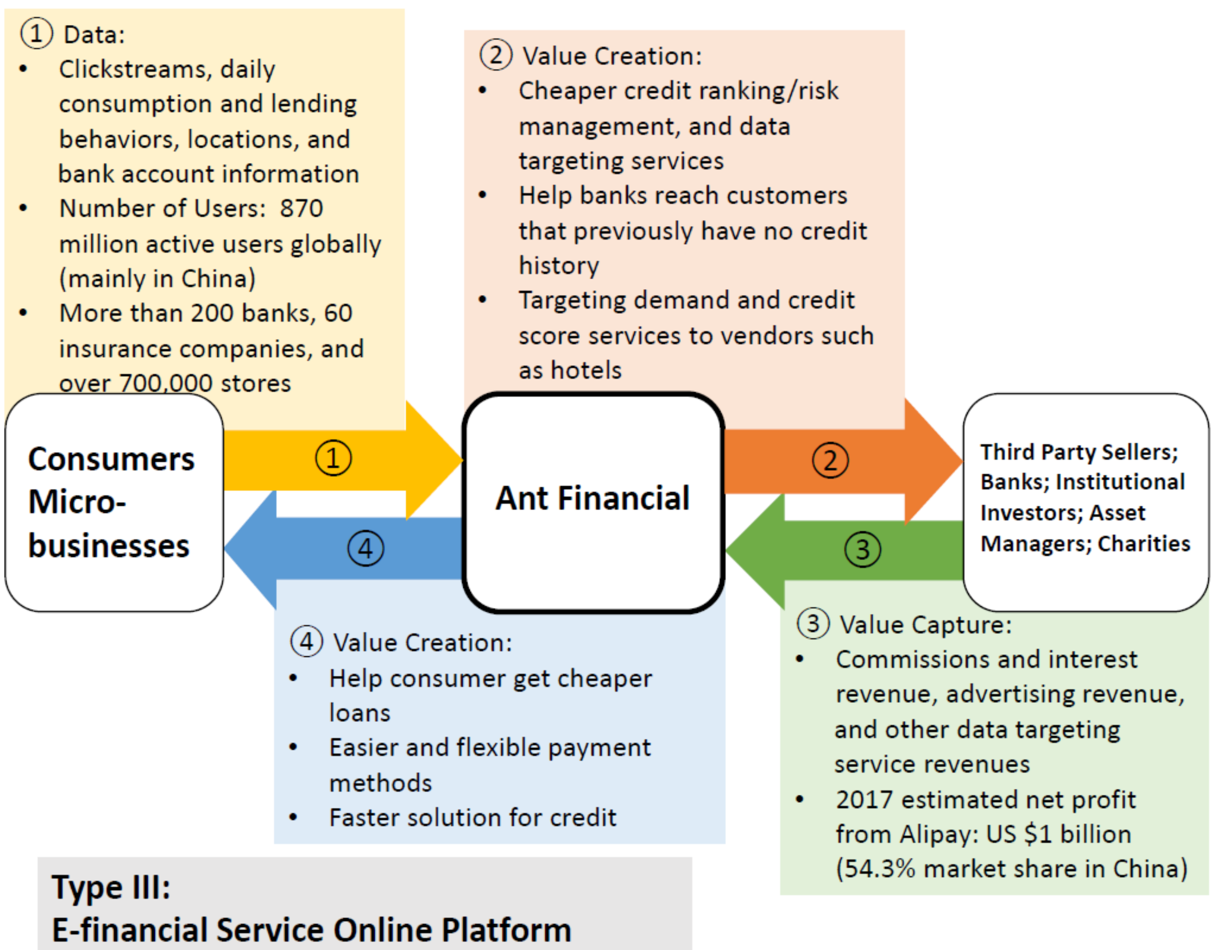


Figure 3: *Type III: E-financial Service Online Platform*

Case Study: Ant Financial

2.2.4 Type IV: Online Social Network Service Platform

Type IV is the social network service online platform, and our case study is LinkedIn (see **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 resumes and to connect with professional friends. The professional network may facilitate their job search. On the other hand, it allows employers to post jobs and to search potential candidates. To date, there are 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 in December 2016, the majority of its revenue came from selling access to its member data, and its revenue in 2015 was US $2.99 billion. In December 2016, Microsoft purchased LinkedIn for US $26.4 billion.

2.2.5 Type V: Online Auction or Matching Platform

Type V is the online auction or matching platform, and our case study is eBay (see **Figure 5**). eBay is a leading online auction platform that facilitates consumer-to-consumer, business-to-consumer, and business-to-business sales.[5] It is free for buyers to use, but sellers are charged fees for listing items after a limited number of free listings and charged again after the items are sold. On the one hand, it provides the buyer with a convenient and cheaper way to purchase products

[5] Some characteristics of eBay are similar to those in the type of e-commerce online platform. However, a separate category is necessary to characterize online platform companies such as eHarmony, which detailed data are not publicly available for inclusion in this study.

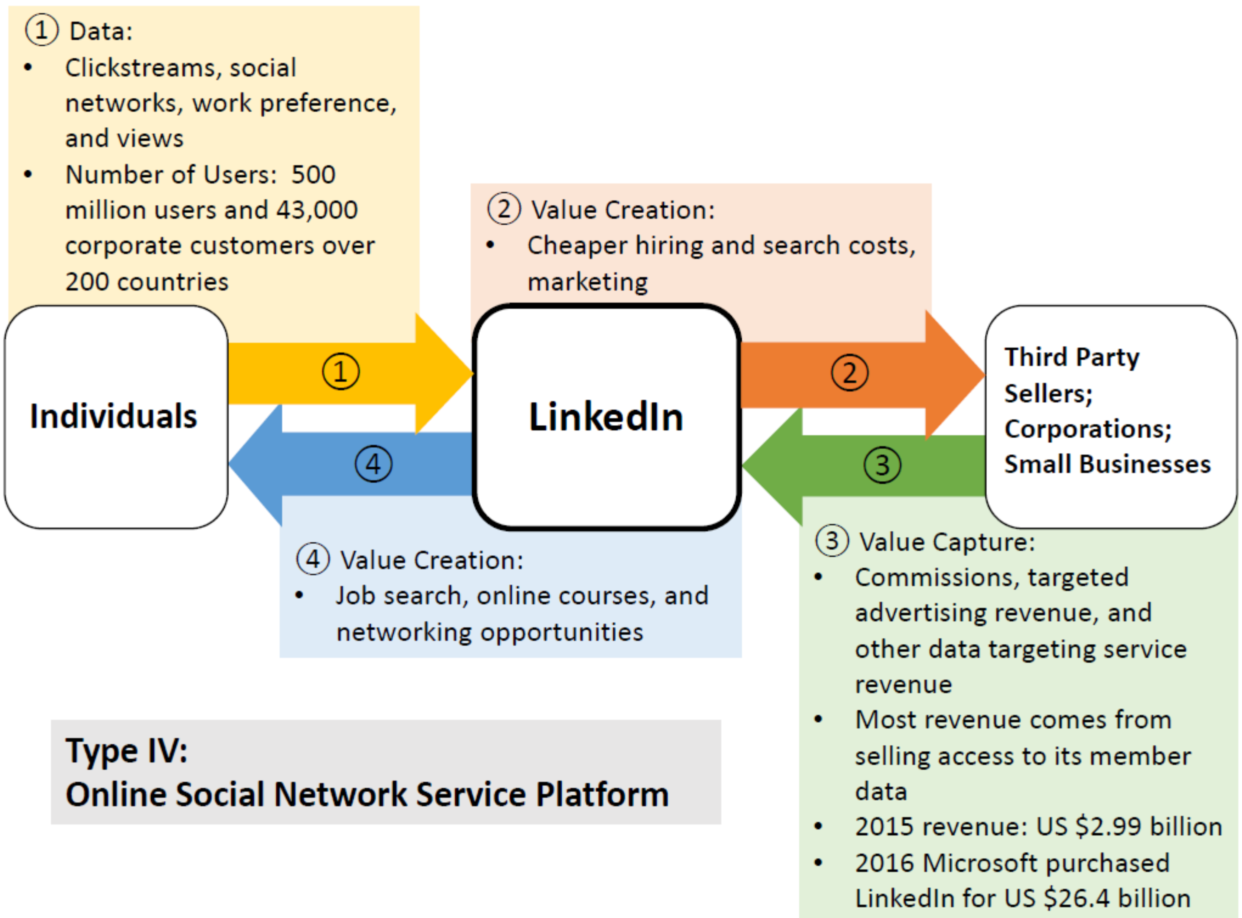


Figure 4: *Type IV: Online Social Network Service Platform*

Case Study: LinkedIn

and/or special collection items. On the other hand, it allows sellers to get access to a big 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 analytics on those data to sell data targeting services. It is reported that eBay has already experienced significant business successes through its data analytics. It currently employs 5,000 data analysts.

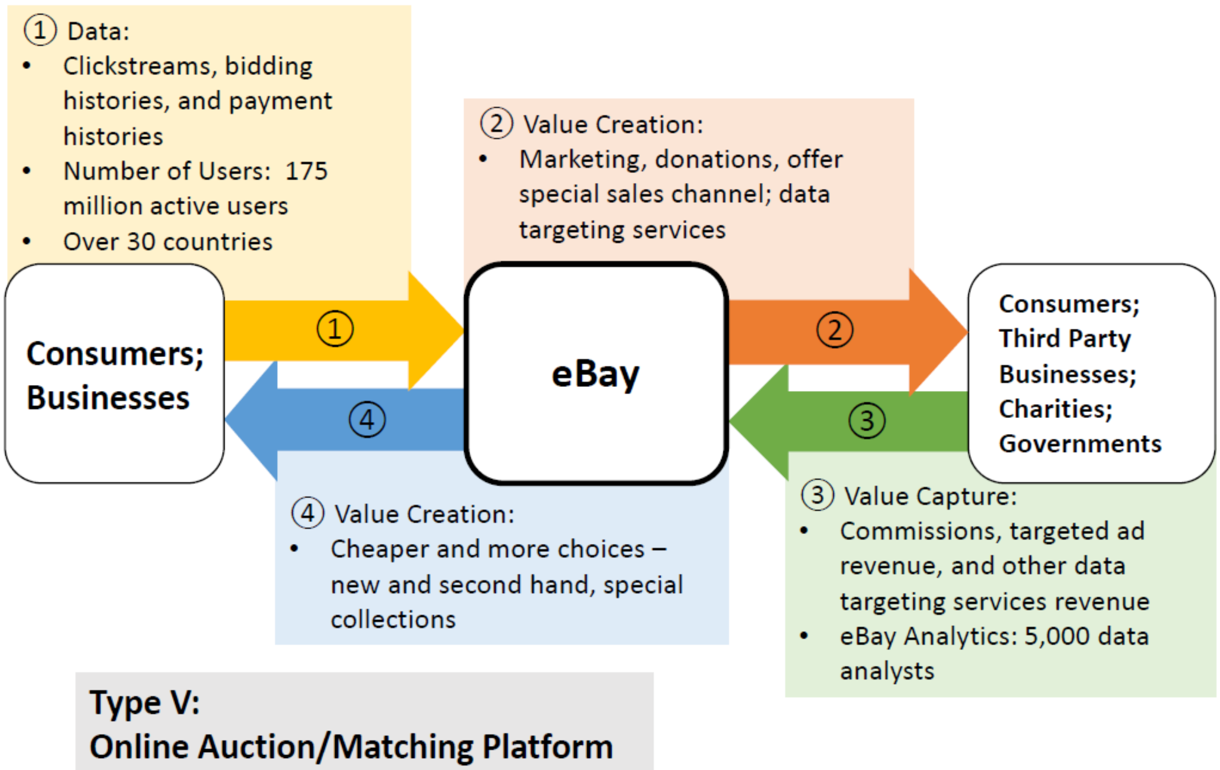


Figure 5: *Type V: Online Auction/Matching Platform*

Case Study: eBay

2.2.6 Type VI: Online Competitive Crowdsourcing Platform

Type VI is the online competitive crowdsourcing platform, and our case study is on Topcoder (see **Figure 6**). An online crowdsourcing platform is a marketplace where an individual or organization can solicit solutions to a certain problem from a large group of outside experts. If the crowdsourcing involves monetary compensation, it is called competitive crowdsourcing. Otherwise, it is called non-competitive crowdsourcing (Chen, 2018).

Topcoder is a popular competitive crowdsourcing platform that organizes the contests on behalf of its clients and sets up the competition rules and rewards. It solicits solutions from its registered members and awards the best solution. The community is formed by designers, developers, data scientists, and competitive programmers. On the one hand, the registered members can get benefits of lower search costs to find suitable contests, learning opportunities through participating in contests, opportunities to conduct challenging work, potentially earning fame such as becoming the “most valuable player,” and possibly earning monetary rewards. On the other hand, it provides a cheaper, faster, and more flexible way for companies to seek solutions. In addition, it allows them to tap into a large group of outside experts but pay only for the best solution. It was reported that researchers from Harvard Medical School, Harvard Business School, and London Business School successfully used the Topcoder Community to solve complex biological problems (Wikipedia, 2018). In a Topcoder challenge to solve a biology-related big-data problem, the best algorithm created by the competitors not only can give more accurate results but is also 1,000 times faster than what was previously available. According to Deloitte (2016), 85% of the top global brands have used crowdsourcing in the past decade.

In terms of the data flow, Topcoder collects data on talents, ideas, and locations from its community. Today, Topcoder has over 1 million registered members. It can use data to solicit businesses and earn commissions based on awards. The clients include big organizations such as NASA, Eli Lilly, Harvard Medical School, and IBM. In 2016, Wipro, one of the world’s largest outsourcing firms, purchased Topcoder for US $500 million.

2.2.7 Type VII: Online Noncompetitive Crowdsourcing Platform

Type VII is the noncompetitive crowdsourcing platform, and our case study is Waze (see **Figure 7**). Waze is a popular noncompetitive 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.

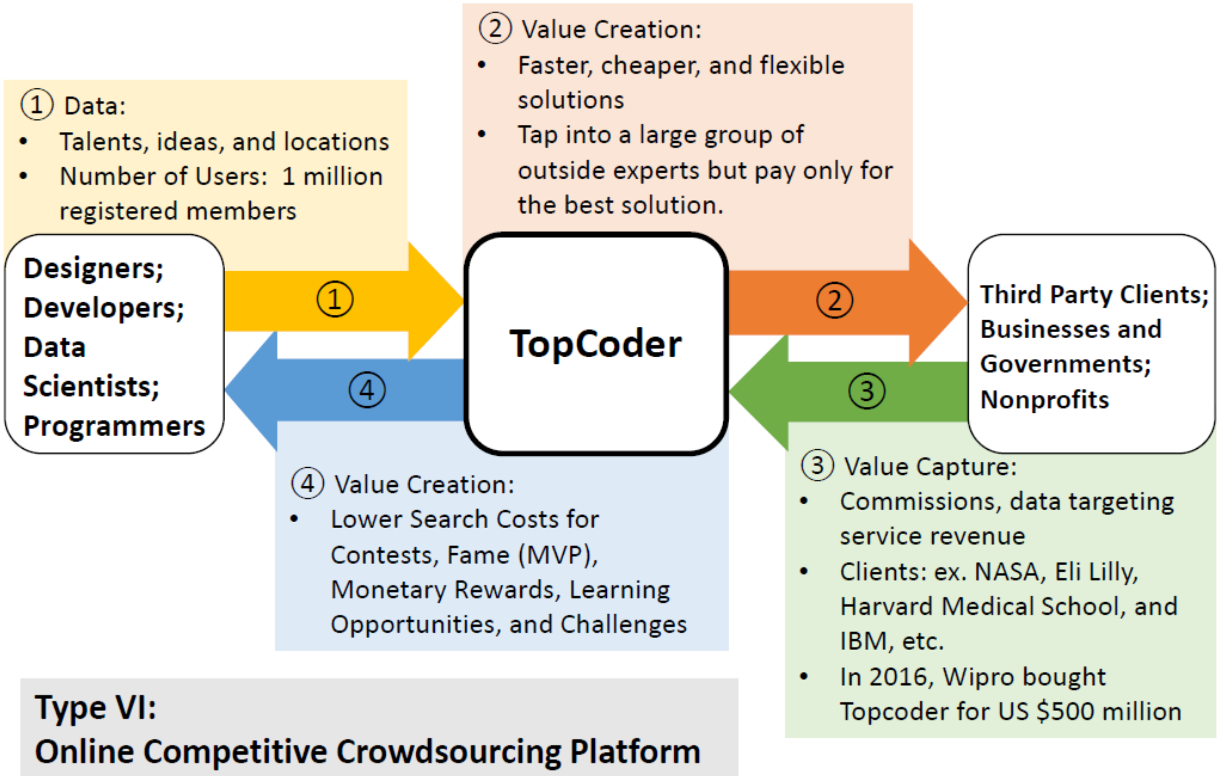


Figure 6: *Type VI: Online Competitive Crowdsourcing Platform*

Case Study: Topcoder

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

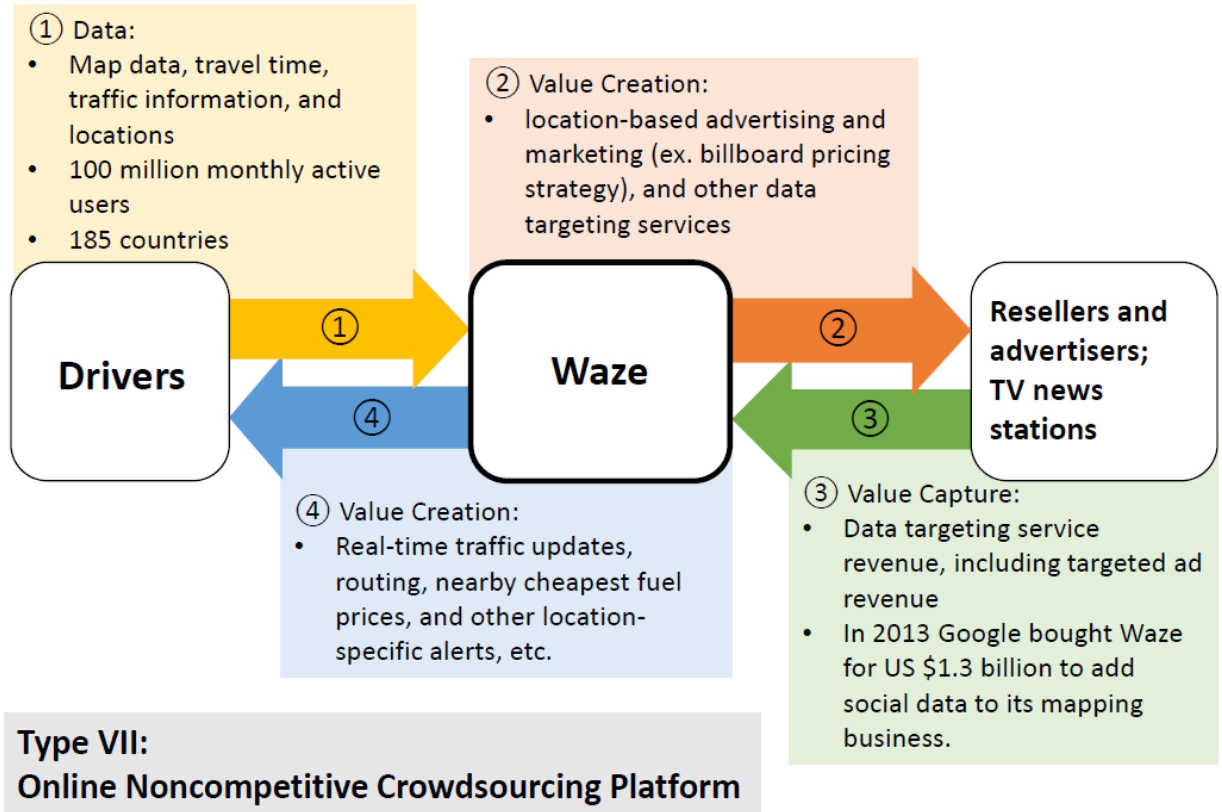


Figure 7: *Type VII: Online Noncompetitive Crowdsourcing Platform*

Case Study: Waze

2.2.8 Type VIII: Online Search Platform

Type VIII is the online search platform, and our case study is Google Search (see **Figure 8**). Google Search is the most popular online platform in the world. On the one hand, it provides individuals a free, convenient, and relevant way to get information instantly. On the other hand, it allows advertisers and content providers to reach a large user base. In addition, the data targeting ads can increase advertisers' returns on investments (ROI). It also allows content providers to add search functionality to their webpages and monetize their content. Given the high degree of integration of Google Play with other Google services, especially Google Search, content is likely to be more easily found with direct links to Google Play, thereby driving more transactions.

In terms of data flow, Google Search collects data on search terms, revealed preferences, browsing behaviors, locations, demographics, languages, etc., from users. It then conducts analytics on those data to provide its corporate clients data targeting services, such as targeted advertising and better demand forecast or marketing. Currently, most of its revenues are from the data targeting advertising revenue. For example, in 2017, Booking Holdings paid Google US $3 billion for AdWords advertising, a part of Google's advertising revenue of US $95.4 billion.

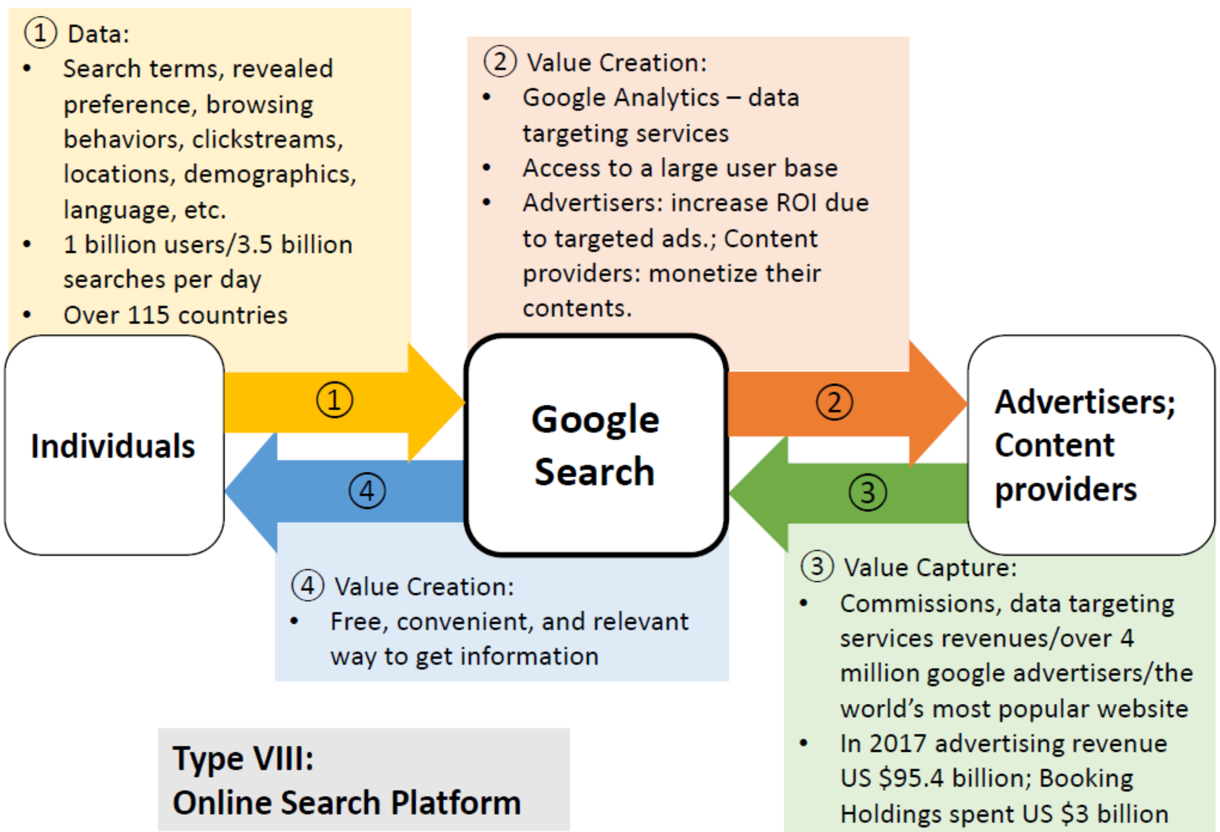


Figure 8: *Type VIII: Online Search Platform*

Case Study: Google Search

3. Data Value Chain and Measurement Issues in Valuing Data

From Section 2, we understand the basic business model underlying in each type of online platforms and data activities associated with each of them, including the data flow, and how an online platform company monetizes its data. Data, a type of intangibles, do not wear and tear. Moreover, unlike regular intangibles, such as R&D assets, which may depreciate due to obsolescence and market competition (Li and Hall, 2018), the aggregation and recombination of data can create new value. That is, new value can be created through ways such as data fusion and/or a creation of new data-driven business models.

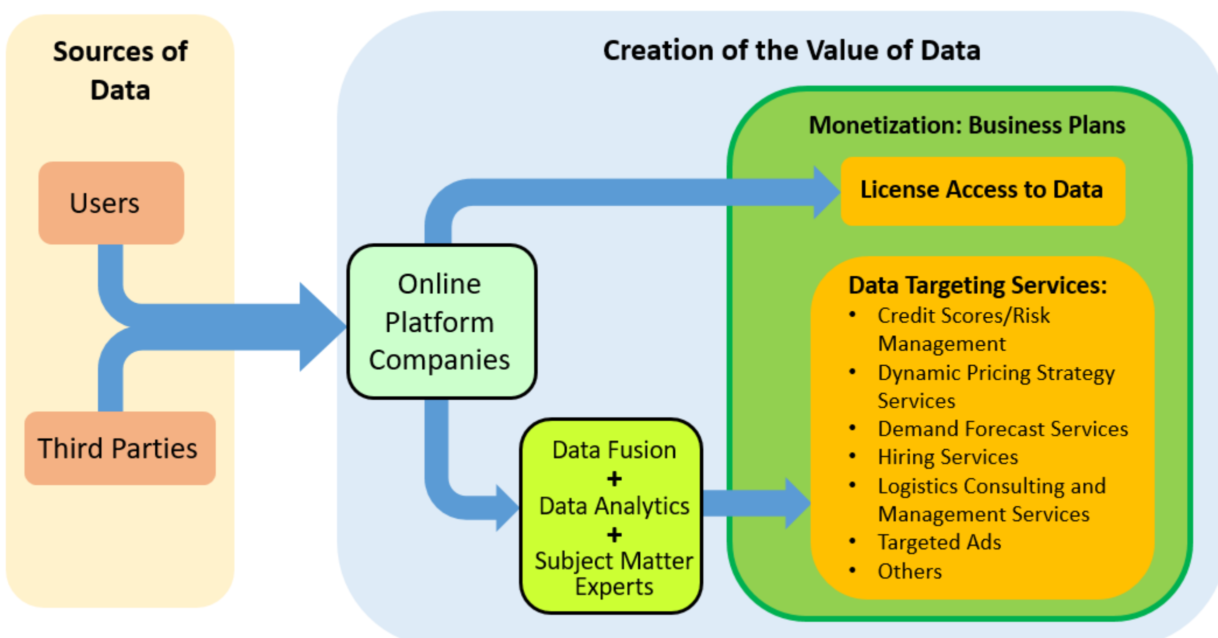
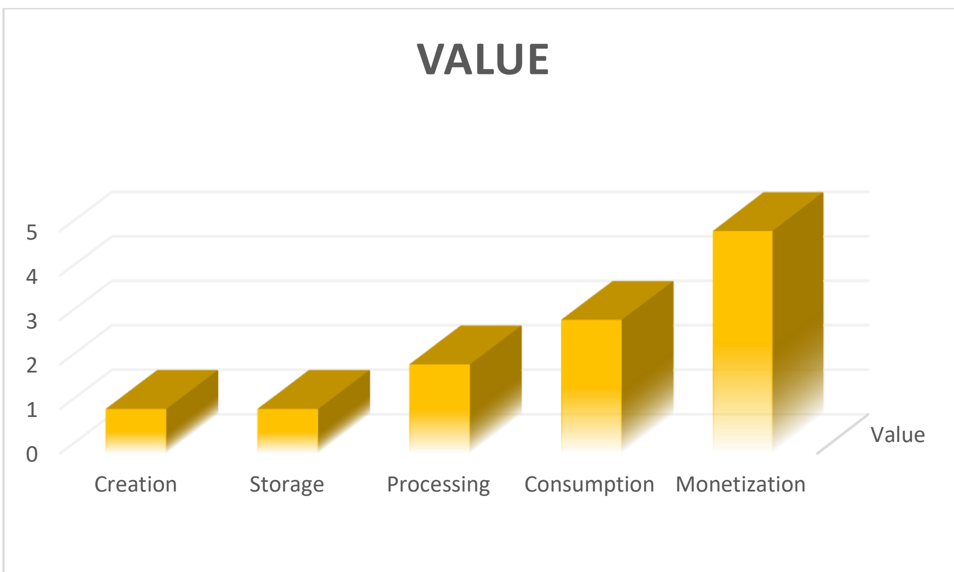


Figure 9: *Creation of the Value of Data*

Figure 9 illustrates how the value of data is created by online platform companies based on our understanding derived from Section 2. In general, 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 the full potential of data, firms such as Twitter may tend to license access to their data to outside companies. 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. From the chart in Figure 9, we can see that the value of data is created through either sales of access to data or sales of data targeting services. Depending on the combined capabilities of their technical skills on data fusion and analytics and business experts, online platform companies can offer a variety of data targeting services that produce revenues. Even within the same type of online platforms, companies can vary by the data targeting services they offer.

Since new values of data can be re-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 re-created through innovations in data-driven business models. These unique features of data pose measurement challenges to capture the value of data. For example, using cost-based approach, such as using the salaries of data analysts and/or data centers as a measure of the value of data, is likely to significantly underestimate the value. Because of the cheap cloud computing services offered by companies like Microsoft and Amazon, many online platform companies outsource their data centers, as discussed previously in the case study on Booking Holdings. The costs of collecting and storing data through data centers are relatively small as shown in the case of Booking Holdings, which is a asset-light but highly profitable company and has a gross profit margin of 98% in 2017.

Figure 9 also shows steps by which firms create the value of data, and how the value varies by step is illustrated in Figure 10. Visconti et al. (2017) define the data value chain by the following five stages: data creation and collection, data storage, data processing (data fusion and analytics), consumption (data visualization and sharing), and monetization (business model). They also give a value on a scale of one to five to each stage, representing the relative amount of the value of data created in each stage. Note that these relative values are not meant to be exact measurements, but they are still useful in understanding how each stage contributes to the creation of the value of data. In the stages of data collection and storage, the value of data is a mere one. The value of data can increase to two in the stage of data processing and to three in the stage of data visualization and sharing. In the final stage of monetization, a data-driven business model under an online platform can create much more value of data. In other words, the data itself may not contain much value, and the majority of the value is generated when a firm has a data-driven business model, which is a business plan that contains monetization-driven organizational planning and cash flow forecasts.



Source: Visconti et al. (2017)

Figure 10: *Data Value Chain*

We can use two examples to explain the concept of data value chain and vertical integration within the chain. The first example is ITA Software-Farecast-Google Flights described in Section

1. Before being purchased by Google, ITA Software focused its business on the first two stages of the data value chain and licensed the use of its data to companies such as Farecast. Farecast focused its business on the third to fifth 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 does not have sufficient in-house data analytics capabilities and/or the capabilities to fully vertically integrate in the data value chain. Even though it had a tremendous amount of data, it chose to sell access to its user data to third-party analytics companies before 2010. Later, when data targeted advertising services became a popular monetization strategy, Twitter also adopted this business model to monetize its data in 2010. In the first quarter of 2018, 12.3% of Twitter's sales comes from selling access to its user data, and the rest of its revenue comes from data targeted advertising revenue, another way to sell the data in the form of advertisement. In addition, Bary (2018) reported that the growth of Twitter's earnings relied on selling its user data, which is one of the two main drivers behind its solid growth in the first quarter of 2018 and reached US $90 million with a high margin and fast growth. In fact, Twitter has four companies as its "official data resellers" that have direct access to all tweets data (Beers, 2018). An understanding of the data value chain can help identify the right approach to measure the value of data.

Three of the existing approaches can be useful in measuring the value of data: the cost-based approach, the market-based approach (Slotin, 2018), and the income-based approach. As mentioned earlier, due to the unique feature of how the value of data is created, a cost-based approach can seriously underestimate the value of data. On the other hand, Akerd and Samani (2018) point out that the assumption that 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 during a merger or acquisition. That is, a firm's inability to evaluate the value of data can result in mispriced products. Moreover, it is impossible to visualize all the possible ways that we can employ the data in the future, especially when significant value can be created through data fusion. All the above considerations imply that both the market-based and income-based approaches are better options even though they may still underestimate the value of data. In this study, we consider new and more rigorous approaches of estimating the value of data, and the methodology is described in the next Section.

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

As shown in Section 2, the underlying business models of online platform companies determine what data they collect, how data flow, and what value of data they create. This part of investments heavily relied on online platform companies' investments in business models, which can be measured by their investments in organizational 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 research, we use the sales, general, and administrative (SG&A) expense as a proxy for a firm's investment in organizational capital (Lev and Radhakrishnan, 2005; Eisfeldt and Papanikolaou, 2013; Brynjolfsson et al., 2018). Firms report this expense in their annual income statements. It includes most of the expenditures that generate organizational capital, such as employee training costs, brand enhancement activities, consulting fees, and the installation and management costs of supply chains, and so on. Because SG&A expenditures may include some items that are unrelated to improving a firm's organizational efficiency, people might question whether it is a valid measure of a firm's investment in organizational capital. Eisfeldt and Papanikolaou (2013) use five ways to validate their measure, and the results show that four out of five ways clearly support this approach.

Moreover, the inefficiency of the investment in organizational capital by definition should show in the depreciation rate of organizational capital. That is, if a firm's investment in organizational capital has a lot of inefficiency, the value of its organizational capital cannot be maintained well, which implies that it will have a higher depreciation rate of organizational capital. As shown in Li (2015), across U.S. high-tech industries, market leaders in general have a smaller depreciation rate than their followers. In this research, we adopt the R&D depreciation model, a forward looking profit model, developed by Li and Hall (2018) to estimate the depreciation rates of the organizational capital for four online platform companies, including Amazon, Booking Holdings, eBay, and Google, for which public data are available. Following Hall (1993), we use

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%	
Online Resource Sharing	Booking	US $11.8 billion (2017)	US $15.7 billion; Annual Growth Rate: 40%	
E-financial Service	Ant Financial	No public financial statement.		
Social Network Service	LinkedIn	US $2.99 billion (2015)**		US $26.4 billion by Microsoft in 2016
Auction/Matching	eBay		US $16 billion; Annual Growth Rate: 30%	
Competitive Crowdsourcing	Topcoder	No public financial statement.		US $500 million by Wipro in 2016
Non-competitive Crowdsourcing	Waze	No public financial statement.		US $1.3 billion by Google in 2013
Search	Google	US $95.4 billion (2017)***	US $48.2 billion; Annual Growth Rate: 21.8%	

* 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.

*** Data targeting service revenue: Data targeted advertising revenue the perpetual inventory method to construct the stocks of organizational 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 based on this approach (see column 4th), annual commission or licensing revenue, and merger & acquisition prices associated with our case studies. For example, 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. These are estimates based on Amazon's financial statements.

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	Home security	USA	Smart Home
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	E-commerce 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	Electronic 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 manufacturer	USA	Data Device
2013	Waze	US $1.3 billion	0.0044	GPS navigation software	Israel	Data
2014	Nest Labs	US $3.2 billion	0.0077	Home automation	USA	IoT; Data
2018	HTC	US $1.1 billion	0.0015	Talent and intellectual property licenses	Taiwan	Data Device

Tables 2 to 4 list the merger and acquisition (M&A) history of Amazon, eBay, and Google, respectively. These tables show that the purpose in the majority of the M&A cases is related to data, indicating that these online platform companies are aggressively expanding the types of data in their collections. In addition, the purchased prices of those M&A cases can provide an indication on how those online platform companies value the data owned by the acquired firms.

Lastly, we have also used a difference-in-difference method and the state space model to assess the causality between the M&A and the stock prices of the acquiring firm in each of these cases (Varian, 2014; Scott and Varian, 2014; Brodersen et al., 2015). We do not find any statistically significant causality effect, and the reason may be due to the size of the deal being too small to affect the market cap of the acquiring firm. More research on alternative methods to identify the causality effect and the change of stock prices will be explored in the future.

5. Discussion on Welfare

Many online platforms have been providing digital goods and/or services to consumers at zero monetary cost in exchange for consumers' data. As large data holders, online platform companies can monetize data by conducting analytics on the data to provide data targeting services and/or licensing the use of the data to third parties. Welfare analysis on digital goods and services without considering the value of data can mislead policy analysis.

For example, transactions through an e-commerce online platform can generate a tremendous amount 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 types of data. 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, as shown in the case study on Booking.com and Marriott International (see Section 2.2.2).

The economic value and welfare implications of those transaction data can be discussed in two scenarios, depending on whether or not the identity of a consumer who engages in the transaction is disclosed. The first scenario considers the condition where the correspondence between a service provider and a consumer is established may cause potential welfare loss such as identity theft or privacy breach as reviewed in Acquisti et al. (2016), where they provide a summary of event-study estimations of the economic impact of data security breaches.

A more subtle effect is dynamic price discrimination. For example, online platform companies can provide data targeting pricing strategy services for third-party sellers. Based on a consumer's data on past 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 he or she would accept the higher price. An extensive review by Acquisti et al. (2016) concludes that "the evidence of systematic and diffuse individual online price discrimination is, currently, scarce." It is reported, however, Booking.com can on average increase the revenue of its corporate clients by 7% through the data-driven pricing strategy service provided to its third-party sellers.

Price discrimination does not necessarily imply 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 its 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 identity of an observed consumer who engages in the transaction is not revealed to or used by the service provider. For example, a consumer can enjoy the benefits gained by revealing his or her attributes, but the provider of the goods or services 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 his or her attributes and transaction records as an observation in data. Namely, the marginal cost of data provision is zero in the case where anonymity is preserved.

A single data point that a data subject provides under anonymity has little value. However, a collection of observations generates a value that cannot be matched by a single data point. 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 does not have a 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 goods or service that fit his or her 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 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. 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 as shown in the Booking case. Moreover, from the perspective of an online platform company, the accumulation of transaction data can increase the productivity of its matching algorithm and the increased productivity through data accumulation is a byproduct of business operation. However, data is certainly an asset, given that online platform companies have earned significant revenues and profits through monetizing data.

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 will enhance the prediction power on both attributes. Thus, as far 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 will 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 emergence of new online platform business models may cause business-stealing effects and negatively affect incumbent firms, 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 emergence of new online platform business models speeds up an economic 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 it does not necessarily imply welfare loss. For example, an online platform company can accumulate a tremendous amount of data that provides a great competitive advantage and render the business model of its conventional competitors obsolete. Consequently, the consumer surplus, income and rent generated by the conventional business decline, and the consumer surplus, income and rent generated by new business increase. Li, Nirei, and Yamana (2018) estimate a lower bound of the redistributed value by conducting case studies on the gains and loss of firms' market valuations.

Lastly, data are information goods and act like knowledge. Therefore, many arguments on knowledge can apply to data as well. 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, Jones and Tonetti (2018) argue that, because data is non-rival, there are potentially large gains by sharing data, the information externality effect of data.

6. Conclusion

Online platform companies are data companies. They provide “free” digital goods and services to consumers in exchange for a tremendous amount of data from consumers. They are normally asset-light but can be extremely profitable. 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 the data, and what consumers get from exchanging their data. In this research, based on the underlying business models, we classify online platforms into eight major types. To understand data activities in each type of online platforms, we conduct case studies to analyze each type of online platforms from the dimensions of business models, data flow, value creation for consumers, value creation for third-party sellers, and monetization of data. Then, based on the case studies for the eight types of online platforms, 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 how the value of data varies by step.

We find that online platform companies can differ in terms of their degrees of vertical integration in the data value chain, which determines how they monetize their data. For example, unlike Google and Amazon, Twitter has a much lower degree of vertical integration in the data value chain and reportedly has 12.3% of its revenue from licensing the use of its user data to data analytics firms (Bary, 2018). That is, online platform companies can monetize their data by licensing the use of data and/or providing data targeting services, but how much economic benefits an online platform company can capture may depend on its degree of vertical integration in the data value chain. As shown in Section 3, online platform companies with business plans to monetize their data 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 is one of the few critical ingredients for these companies to create data-driven digital goods and services (Brynjolfsson et al., 2018; Lee, 2018). With current cheap computing power and an adequate supply of AI engineers, online platform companies with a higher degree of vertical integration in the data value chain can benefit more from data. Furthermore, their businesses can be strengthened by the virtuous cycle between deep learning's relationship with data described by 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.

Data, a type of intangibles, do not wear and tear. Moreover, unlike regular intangibles, such as R&D assets, which may depreciate due to obsolescence, data can produce new values through the aggregation and recombination of data. These unique features of data pose significant challenges to firms and statistical agencies in measuring the value of data. In this study, we propose a way to estimate the value of data for the representative firms in the eight types of online platforms. The initial results indicate that the value of data can be significant: For example, the value of Amazon's data can account for 16% of Amazon's market valuation and has an annual growth rate of 35%. However, because some online platform companies are private, we do not have the data to perform the same estimation for all the firms in our case studies, which is an area left for future work. In addition, the current depreciation model assumes decreasing marginal returns to data-driven business model investments. Future research could modify the model to incorporate the increasing marginal returns to the investments. Moreover, we review the guidelines of the 2008 System of National Accounts in Appendix B and discuss how the guidelines can apply to the measurement of the value of data.

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 strategies for its corporate clients, such as price discrimination strategies based on consumers' data on clickstreams and past transactions, to maximize their revenues, as shown in the Booking case. On the other hand, the households in the economy as a whole can receive the equivalent value of additional income. In addition, the price discrimination can be redistributive. 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 price discrimination does not necessarily lead to a more equal distribution. 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. A single data point that a data subject provides under anonymity has little value. However, a collection of observations generates a value that cannot be matched by a single data point. 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 does not have a value, but a collection of them potentially does. Therefore, 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 goods or service that fit his or her 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 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. 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 as shown in the Booking case.

Online platforms are evolving rapidly. To seek growth, online platform companies may expand their existing business models from covering one type of online platform to multiple types. This is similar to the situation where a firm conducts businesses in multiple industries. That is, we may see online platform companies develop hybrid online platforms that cover several basic types of online platforms that we classify below. Also, the degree of hybrid platforms can vary across countries. As explained by Lee (2018), a few Chinese online platform companies have developed super online platforms where each of them, such as Tencent's WeChat, bundles many online platform functionalities similar to Facebook, Uber, Expedia, PayPel, Amazon, LimeBike, and more combined, an outcome that can be called as "The App Constellation Model." In contrast, most U.S. online platforms are less hybrid and focus on original business models. More research is needed to understand the impacts of the rapidly evolving trend of online platforms in areas including data collection, market competition, and consumer welfare.

Lastly, data is like new oil. At present, the pipelines of new oil are controlled by online platform companies. In the future, blockchain technology may allow each consumer to have his or her own pipeline, to take control of their ownership, and to decide whether and how to sell personal data to AI companies, online platform companies or advertising companies. However, given the fact that the creation of the value of data takes place on the company side and depends on data analytics and business experts to build a business plan to monetize data, consumers are not in a position to understand the value of their personal data. Nevertheless, how fast various industries can adopt blockchain technology may affect the future competition of online platform companies. At present, data volume doubles every three years (Mayer-Schonberger and Cukier, 2014), but the Internet of Things (IoT), the trend of 5G, and the emerging online-to-offline transition are rapidly accelerating the accumulation speed of data types and volume. Therefore, it is very important to develop feasible methodologies to measure the value of data. The valuation of data is not only important 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, how to monetize them, 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 are different in terms of the ownership of personal data, for example, Europe's new data protection rule, the General Data Protection Regulation, and China's extreme openness of personal data. Additionally, the U.S. allows foreign firms to collect personal data in the U.S. but China prohibits it. How do the differences in data policy affect trade? Given the existence of the virtuous cycle between deep learning'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 trade and growth.

References

Acquisti, A., Taylor, C., and 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.

Bary, E. 2018. Twitter Earnings Growth Relied on Selling User Data, *Market Watch*, April 25th. https://www.marketwatch.com/story/twitter-earnings-growth-relied-on-selling-user-data-2018-04-25

Beers, B. 2018. How Does Twitter Make Money? Investopedia, July 27th. https://www.investopedia.com/ask/answers/120114/how-does-twitter-twtr-make-money.asp

Bloomberg. 2018. Facebook Cambridge Analytica Scandal: 10 Questions Answered, *Fortune*, April 10th, http://fortune.com/2018/04/10/facebook-cambridge-analytica-what-happened/

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., Eggers, F., & Gannamaneni, A. 2018. New Measures of the Economy: Measuring Welfare with Massive Online Choice Experiments: A Brief Introduction, *AEA Papers and Proceedings*, 108: 473-476.

Brynjolfsson, E., Hui, X., Liu, M. 2018. Does Machine Translation Affect International Trade? Evidence from A Large Digital Platform, *National Bureau of Economic Research Working Paper No. 24917*, August.

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.

Chen, L.Y. 2018. Ant Financial Annual Profit Jumps 65% Ahead of Anticipated IPO, *Bloomberg*, May 4.

Chen, Y., Dai, T., Korpeoglu, C., Korpeoglu, E., Sahin, O., Tang, C., & Xiao, S. 2018. Innovative Online Platforms: Research Opportunities, working paper.

Cohan, P. 2013. Google To Spite Facebook, Buy Wave For $1.3 Billion, *Forbes*, June 9th, https://www.forbes.com/sites/petercohan/2013/06/09/google-to-spite-facebook-buy-waze-for-1-3-billion/#4e0ef10d6228.

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.

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, https://www.bloomberg.com/news/articles/2018-08-08/is-apple-really-your-privacy-hero

Hall, B.H. 1993. The Stock Market's Valuation of R&D Investment During the 1980's, *The 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, https://www.businessinsider.com/airbnb-total-worldwide-listings-2017-8

Hathaway, I., & Muro, M. 2017. Ridesharing Hits Hyper-growth, https://www.brookings.edu/blog/the-avenue/2017/06/01/ridesharing-hits-hyper-growth/

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*.

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, *the Review of Income and Wealth*, in press.

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

Molla, R. 2017. Amazon Could Be Responsible for Nearly Half of U.S. E-commerce Sales in 2017, *Recode*, October 24th, https://www.recode.net/2017/10/24/16534100/amazon-market-share-ebay-walmart-apple-ecommerce-sales-2017.

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. https://www.sec.gov/Archives/edgar/data/1326801/000132680113000003/fb-12312012x10k.htm

SEC. 2017. Booking Holding Inc. (BKNG) SEC Filing 10-K Annual Report for the Fiscal Year Ended December 31, 2017, https://www.last10k.com/sec-filings/bkng

Shu, C. 2015. Data From Alibaba's E-commerce Sites Is Now Powering A Credit-Scoring Service, *techcruch.com*, Jan. 28th.

Slotin, J. 2018. What Do We Know About the Value of Data? May 3rd. http://www.data4sdgs.org/news/what-do-we-know-about-value-data

Statista. 2018. Percentage of Paid Units Sold by Third-party Sellers on Amazon Platform As of 2nd Quarter 2018, https://www.statista.com/statistics/259782/third-party-seller-share-of-amazon-platform/

Tefis Team. 2017. Facebook's Strong Ad Revenue Growth To Continue, *Forbes*, October 30th, https://www.forbes.com/sites/greatspeculations/2017/10/30/facebooks-strong-ad-revenue-growth-to-continue/#254871fe6fe7

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, Chapter in forthcoming NBER book, *The Economics of Artificial Intelligence: An Agenda*, edited by Joshua Gans and Avi Goldfarb. The University of Chicago Press.

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.

Weinland D., and Ju, S.F. 2018. China's Ant Financial shows cashless is king. *Financial Times*, April 13.

Whitmore, G. 2018. How Google Travel Is Changing the Game with Google Flights, Google Trips, and More, Forbes, February 20th. https://www.forbes.com/sites/geoffwhitmore/2018/

02/20/how-google-travel-is-changing-the-game-with-google-flights-google-trips-more/#7683e1b34d98.

Wikipedia. 2018. Topcoder, https://en.wikipedia.org/wiki/Topcoder

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

Appendix A: Depreciation Model of Business R&D Capital

Li and Hall (2018) develop a depreciation model to estimate U.S. business R&D capital. The premise of the model is that R&D capital depreciates because its contribution to a firm's profit declines over time and the main driving forces for the decline are the pace of technological progress and the degree of industry competition. The model can apply to other intangible capitals as long as the data on investments and sales are available. Below is the brief summary of the model.

A profit-maximizing firm will invest in R&D such that the expected marginal benefit equals the marginal cost. That is, in each period *t*, a firm will choose an R&D investment amount to maximize the net present value of the expected returns to R&D investment:

$$\max_{R_t} E_t[\pi_t] = -R_t + E_t\left[\sum_{j=0}^{\infty} \frac{q_{t+j+d} I(R_t)(1-\delta)^j}{(1+r)^{j+d}}\right] \qquad (1)$$

where R_t is the R&D investment amount in period *t*, q_t is the sales in period *t*, $I(R_t)$ is the profit rate due to R&D investment, δ is the R&D depreciation rate, and *r* is the cost of capital. The parameter *d* is the gestation lag and is assumed to be an integer which is no less than 0. R&D investment in period *t* will contribute to the profits in later periods but at a geometrically declining rate. They assume that the sales *q* for periods later than *t* grows at a constant growth rate, g. That is, $q_{t+j} = q_t\left(1+g\right)^j$. This assumption is consistent with the fact that the U.S. output of most R&D intensive industries grows fairly smoothly over time.

To resolve the issue that the prices of most R&D assets are generally unobservable, the study defines *I(R)* as a concave function:

$$I(R) = I_{\Omega}\left(1-\exp\left[\frac{-R}{\theta}\right]\right) \qquad (2)$$

with $I''(R) < 0.$, $I'(R) = \frac{I_\Omega}{\theta}\exp(\frac{-R}{\theta}) > 0$, $I'(0) = \frac{I_\Omega}{\theta}$, and $\lim_{R\to\infty} I(R) = I_\Omega$. This functional form has few parameters but nevertheless shows the desired concavity with respect to *R*. In this, the approach is similar to that adopted by Cohen and Klepper (1996), who show that when there are fixed costs to an R&D program and firms have multiple projects, the resulting R&D productivity will be heterogeneous across firms and self-selection will ensure that the observed productivity of R&D will vary negatively with firm size. The model incorporates the assumption of diminishing marginal returns to R&D investment implied by their assumptions, which is more realistic than the traditional assumption of constant returns to scale (Griliches, 1996). In addition, the model implicitly assumes that innovation is incremental, which is appropriate for industry aggregate R&D, most of which is performed by large established firms.

The function *I* includes a parameter θ that defines the investment scale for increases in R&D and acts as a deflator to capture the increasing time trend of R&D investment as a component of investment in many industries. The value of θ can vary from industry to industry, allowing different R&D investment scales for different industries.

Using this function for the profitability of R&D, the R&D investment model becomes the following:

$$\begin{aligned} E_t[\pi_t] &= -R_t + E_t\left[\sum_{j=0}^{\infty}\frac{q_{t+j+d}I(R_t)(1-\delta)^j}{(1+r)^{j+d}}\right] \\ &= -R_t + I_\Omega\left[1-\exp\left(-\frac{R_t}{\theta_t}\right)\right]\sum_{j=0}^{\infty}\frac{E_t[q_{t+j+d}](1-\delta)^j}{(1+r)^{j+d}} \end{aligned} \qquad (3)$$

Note that they have assumed that *d*, *r*, and δ are known to the firm at time *t*. Because θ varies over time, they model the time-dependent feature of θ by $\theta_t \equiv \theta_0(1+G)^t$, where *G* is the growth rate of θ_t. To estimate G, they assume that the growth pattern of industry's R&D investment and its

R&D investment scale are similar and they estimate G by fitting the data for R&D investment to the equation, $R_t = R_0 \left(1+G\right)^t$. Using this assumption, Equation (3) becomes:

$$\pi_t = -R_t + I_\Omega \left[1 - \exp\left(-\frac{R_t}{\theta_0 (1+G)^t}\right)\right] \frac{q_t \left(1+g\right)^d}{\left(1+r\right)^{d-1}\left(r+\delta-g+g\delta\right)} \qquad (4)$$

Note that because of their assumptions of constant growth in sales and R&D, there is no longer any role for uncertainty in this equation, and therefore no error term. Assuming profit maximization, the optimal choice of R_t implies the following first order condition:

$$\frac{\partial \pi_t}{\partial R_t} = \frac{(1+G)^t}{I_\Omega} \theta_0 exp\left[\frac{R_t}{\theta_0 (1+G)^t}\right] + \frac{q_t (1+g)^d}{(1+r)^{d-1}(r+\delta-g+g\delta)} = 0 \quad (5)$$

For estimation, they add a disturbance to this equation (reflecting the fact that it will not hold identically for all industries in all years) and then estimate θ_0 and the depreciation rate δ.

Appendix B: The Guideline of the 2008 System of National Accounts on Database and the Measurement of the Value of Data

As described in Section 3, a data value chain has five defined sequential steps: creation (data capture), storage (data warehouse), processing (data mining and fusion), consumption (visualization and sharing), and monetization (business plan). The stage of creation includes both raw data capture and experimental data collection (e.g. A/B testing). The value of data in the stages of creation and storage remains low due to the facts that data can be unprocessed regardless of whether or not the data is structured, and that storage hardly increases the value. Even in those two stages, online platform companies can monetize the raw data by licensing the use of those data, such as in the case study of Twitter.

According to the 2008 System of National Accounts (SNA) recommendations on databases, the cost of preparing data in the appropriate format is included in the cost of the database but not the cost of acquiring or producing the data (SNA, 2008). In these two stages, online platform companies' investments in collecting and storing data, such as data centers, cloud computing, and software, will not count in the calculation of the value of data. In the third stage associated with data processing, the value of data can be increased by data fusion and data analytics. Nonetheless, its output is still not ready for immediate use. This part of investment in the preparation of data should be counted by the investment in software and the costs of data scientists and analysts.

In the last two stages of the data value chain associated with consumption and monetization, the value of data jump to the second highest and highest levels (see Figure 10). From Sections 2 and 3, we understand that online platform companies can monetize their data through two major channels: licensing access to data and creating data-driven business models, i.e., selling the data in different forms of businesses, such as a demand forecast. In paragraph 10.112 of the 2008 SNA, it is stated that "Database may be developed exclusively for own use or for sale as an entity or for sale by means of a license to access the information contained. The standard conditions apply for when an own-use database, a purchased database or the license to access a database constitutes an asset." In addition, in paragraph 10.114 of the 2008 SNA states that "Database for sale should be valued at their market price, which includes the value of the information content." Based on those two guidelines, licensing access to data and data targeting services offered by online platform companies should be included in the value of data.

Moreover, in paragraph 10.113 of the 2008 SNA states that "The creation of a database will generally have to be estimated by a sum-of-costs approach. … Other costs will include …, an estimate of the capital services of the asset used in developing the database, …" As shown in Section 2, the underlying business models of online platform companies determine what data they collect, how data flow, and what value of data they create. This part of investments is not included in the software investment and heavily relies on online platform companies' investments in business models. For example, Booking Holdings' high gross profit margins (see Section 2.2.2) indicate that the company is extremely physical-asset-light but intangible-capital intensive (Li, Nirei and Yamana, 2018).