IARIW 2021

Monday 23 – Friday 27 August

Statistisk sentralbyrå
Statistics Norway

The Data Economy: Market Size and Global Trade

Diane Coyle

(University of Cambridge)

Wendy Li

(Moon Economics Institute)

Paper prepared for the 36th IARIW Virtual General Conference
August 23-27, 2021
Session 23: National Accounts
Time: Thursday, August 26, 2021 [16:30-18:30 CEST]

The Data Economy: Market Size and Global Trade

Diane Coyle[1] and Wendy Li[2]

Running title: The Data Economy

Abstract

Large online platforms using data operate globally. Valuing data is problematic, yet this is crucial for informed policy decisions. We demonstrate a novel sectoral methodology for estimating the economic value of markets for data, estimating the market size for data in the global hospitality industry at US$43.2 billion in 2018. The scale of data flows affects the international division of labor, with important policy implications. With many jurisdictions introducing different data protection and trade regimes, affecting data access by market participants, we present a typology of countries and discuss their ability to benefit from data value creation.

JEL Codes: O3, F1, F2

Key words: data, digital, trade, data centers, platforms

[1] University of Cambridge; dc700@cam.ac.uk. Coyle receives funding from the Economic Statistics Centre of Excellence.

[2] Moon Economics Institute; wendyli@moonecon.org

1. Introduction

Data has an important and growing role in the modern economy, raising questions in several policy domains. It has become an issue in trade, as underlined by announcements such as the latest United States-Mexico-Canada Agreement on digital trade and Japan's announcement on Data Free Flow with Trust for the 2019 G20. Data has also become a key asset for innovation, as documented by many recent studies of online platforms and interviews with executives (eg Niebel et al 2018, Lopez, 2016). On the other hand, the data advantage of online platform companies is increasingly recognized as an entry barrier, leading to concentration in digital markets (eg Furman 2019, Scott Morton 2019). The data gap applies not only at the firm level but also at the country level, and is increasingly leading to debates outside the US and China (where most large technology companies are based) about the need for policy intervention to level the competitive playing field (eg Gerardin 2018). In particular, the gap is growing wider for small, developing countries that lag behind in their digital investments and capabilities. Moreover, countries are forming different data protection and trade regimes, such as those of the EU and Japan and by China. The differences between regimes can affect the data gap and the relative strengths of market participants.

These developments are occurring as the role of data in the economy continues to expand. The era of 5G and the Internet of Things (IoT), the combined forces of artificial intelligence (AI) and big data, are reshaping how and where goods and services are produced and distributed. Online platforms based on AI and troves of data have disrupted each industry sector they have entered, with their firm-specific knowledge of how to use data to produce, compete, and grow. When an online platform enters an industry sector, traditional incumbents with a higher degree of digital capability will adjust better than their counterparts with a lesser degree of digital transformation (Li and Chi, 2021). The COVID-19 pandemic has accelerated the pace of digital disruption in almost every aspect of daily life and business, pushing more traditional firms to accelerate their digital transformation.

Although the production of and demand for data is rising rapidly, most data transfers are unobserved. Firms can trade data through data brokers but most data transfers do not occur this way, and many also involve a data center in a third country. Markets for data (once gathered) are mostly unregulated but have been growing dramatically. How big is the market for data? How fast is it growing in scale and scope? Does the growth vary by industry and by country? These questions are important because their answers are crucial for market participants to make investment decisions, both at firm level and national level. For example, while investments in data infrastructure can be expensive or even infeasible for small developing countries, accessing global data markets via the cloud can lower entry barriers for firms in small developing markets, and can provide them with sufficient incentives to invest locally, given adequate skills. Such decisions will then affect data trade flows.

Understanding the market size for data is therefore critical, but the measurement of market size is challenging because there is no data on most data transfers. There are few market prices, these are not transparent where they exist, and moreover most data are collected for firms' own use, or for exchange with other firms without a monetary transaction. There is a dearth of reliable data on volumes and moreover it is hard to put a monetary value on data even where volume estimates are available (Coyle & Diepeveen 2020). Finally, firms are unwilling to compile or release data on data for reasons of commercial confidentiality.

In this paper, we make two contributions. We develop a novel methodology for measuring the size of markets for data by sector, and the associated annual growth rates, from the demand side. We use the hospitality industry as an example to apply this measurement approach. Our initial estimates indicate that the market for data is substantial and growing fast. Our preliminary, conservative estimate of the market size for data in the global hospitality industry is US $43.2 billion with a growth rate currently doubling market size every three years.

Secondly, we also develop a typology of trade in data and digital goods and services at the country level. We identify six different country categories. As the size and growth of markets for data will differ across industries and countries, both industry-level and country-level market size for data and the associated growth can impact the international division of labor. Population size, availability of high-tech talent, financing availability, digital infrastructure, foreign direct investment, and institutional contexts in different countries will therefore affect trade in data, and in digital goods and services.

The rest of this paper proceeds as follows: Section 2 briefly describes the relevant literature on how digitalization and data use affect firms' organizational capital and production choices, and discusses the implications for trade. Section 3 sets out our data valuation methodology, building on the way data use sustains or enhances firms' organizational capital, along with our empirical results for the hospitality industry. Section 4 presents our typology of trade in data and digital goods and services. Section 5 concludes.

2. Literature

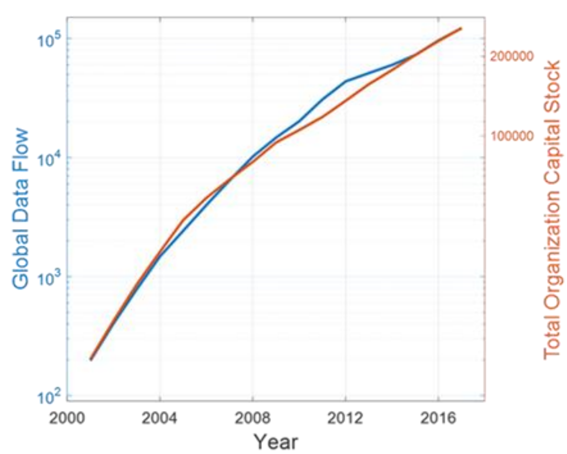
There is an extensive literature on the impact of digitalization on production and consumption, including the rapidly growing use of data. Digital transformation requires firms to reinvent or re-orient their business models. The process is being led by online platforms, which, based on the use of big data and AI, have disrupted incumbent firms' specific knowledge, derived from their relatively limited amount of data, in many sectors. Traditional, non-digitalized firms can see the value of their firm-specific knowledge depreciate faster once disrupted by such entry (Li and Chi, 2021). Knowing how to be productive is the reason why a firm continues to exist (Demsetz, 1997). Hence, firms able to use data to reorganize production enjoy higher productivity (eg e.g. Atkeson and Kehoe (2005), Lev and Radhakrishnan (2005), Brynjolfsson et al (2018) while faster depreciation of an incumbent firm's specific knowledge will lead to lower productivity growth and market valuations (Li and Chi, 2021; Eisfeldt and Papanikolaou, 2013).

Digital transformation and new business models are driving the rapid growth in data use. For example, startups need data to test their algorithms and business models, existing firms need to collaborate with online platforms to get access to their customers' data, online platforms need to trade data for the purpose of diversification, and participants in supply chains need to exchange data to build strategic coalitions. Moreover, thousands of manufacturers around the world have increasingly adopted technologies such as the lights-out factory, digital fabrication, and/or additive manufacturing. Manufacturers are increasingly transforming their business models from selling products to selling services (Tzuo and Weisert, 2018). All these models have a crucial input: data.

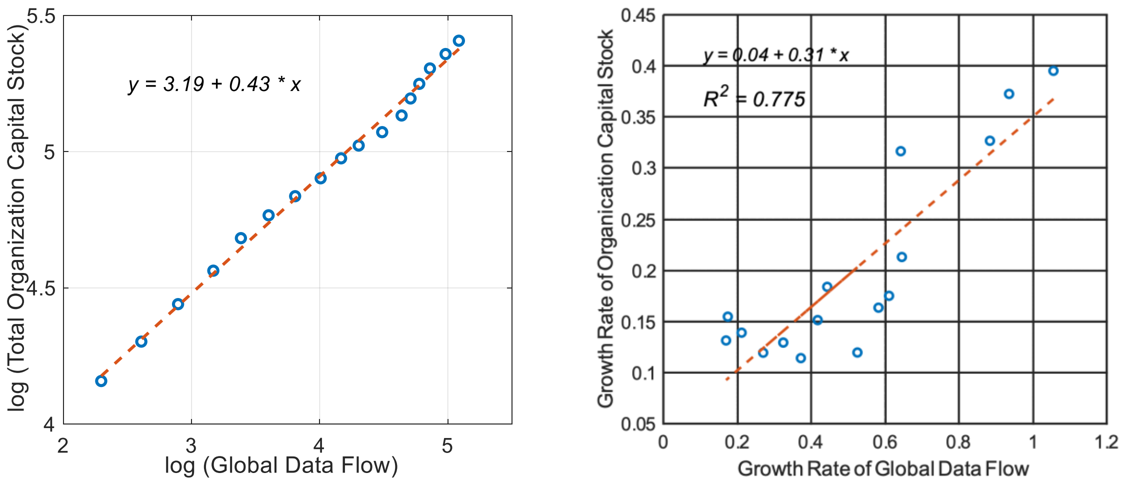
The data input requirement to compete effectively can be substantial, for traditional industry boundaries are blurring. For example, the Apple Watch, launched less than five years ago, already outsells the entire Swiss watch industry with its 152-year history (Thornhill, 2020). AI, using big data sets, is becoming a cheaper and more adaptable tool, as cloud computing companies enable customers to rent their IT needs at much reduced cost and to use advanced AI tools without needing sophisticated in-house skills (Byrne, Corrado & Sichel 2020, Coyle & Nguyen 2019, Marko 2019). It is now possible for a firm to outsource almost everything from data collection to customer relationship management, and from data storage to data analytics, although some firm-specific domain knowledge cannot be outsourced. The data available determine the overall power and accuracy of an algorithm, and thus are vital for firms' competitiveness. For instance, a recent survey of Japanese firms found that firms collecting any data are more productive than those not collecting data, firms collecting more data are more productive, and firms collecting data overseas are more productive than the ones only collecting data domestically (Tomiura et al., 2020).

Such results are in line with Li and Chi (2021)'s finding that data use enables firms to derive firm-specific knowledge, which can be measured by their organizational capital, the accumulated information or know-how of the firm (Prescott & Visscher 1980); the more data, the greater the potential derived firm-specific knowledge. Multinational firms in general collect more cross-border data, and are more productive. Li and Chi (2021) estimated the organizational capital for each of top seven global online platform companies, Microsoft, Amazon, Apple, Google, Facebook, Alibaba, and Tencent, and compared their combined organizational capital with the global data flow during the same period of time (Figure 1). This correlation is suggestive of the fact that large online platform companies have been aggressively investing in specific organizational capital. Li and Chi (2021) also finds that online platforms are more intensive in organizational capital than others. Eisfeldt and Papanikolaou (2013) similarly find that firms are more productive when they are more organizational capital intensive.

Figure 1: Global Data Flow vs. Combined Organizational Capital Stock of Dominant Online Platforms



(1a)



(1b) (1c)

Source: Li and Chi (2021)

Large data holdings, rich in volume and variety, thus give large online platforms a significant competitive advantage, powered by network effects and the virtuous cycle between data and the AI algorithms improving the services and increasing revenues.[1] The data advantage also allows them to gain insights about adjacent sectors and enter them more easily. As online platforms have disrupted the industries they have entered, this has pushed existing firms of all sizes to digitally transform their organizations and business models, driving rapid growth in the demand for data. Potential new competitors without access to data will inevitably struggle to enter the market.

The changing pattern of organizational advantage has implications for the location of production and for trade. Importantly, as data-intensive companies, online platforms can scale up their business operations quickly, not only domestically but also internationally, without facing the traditional physical operational constraints, subject only to cross-border legal constraints on data transfer. The globally dominant online platforms are multinational firms but their operations can be geographically concentrated in a few countries. For example, Booking.com serves 137,791 destinations, with 28.9 million properties in 229 countries (Li et al., 2019). But most of its operations are conducted in its Amsterdam headquarters, where it employs 1,800 engineers accounting for 90% of its workforce (Yin, 2018). Comparing Booking.com and Marriott, Booking has more listed properties and greater scalability (i.e., fewer physical operational constraints), and its 2017 gross profit margin was 98%, compared with the figure of 71.8% for Marriott.

Even for smaller data-intensive firms, the required minimum operational scale is lower than previously and it is possible to scale up operations fairly easily using ever-cheaper cloud computing services (Coyle & Nguyen 2019). Along with (close to) zero transportation costs, this contributes to the concentration of most of the high value-added work in one location. For example, Sweden-based Hermes Medical Solutions with an online platform, GOLD, produces software applications to monitor organ functions of patients. All the data are stored in Sweden, but its cloud-based operations cover more than 40 countries and 95% of its services are for foreign markets (Castro & McQuinn, 2015; Segulah, 2020).

There has been growing concern around the world about the impact of dominant online platforms. Recent reports have identified the data entry barrier and are prompting competition policy changes in some developed country jurisdictions. Baldwin (2019) has argued that what he terms ‘globotics’, the automation of services as well as manufacturing, allowing new patterns of delocalization to emerge, will transform prior patterns of comparative advantage. There are also concerns among developing countries about how they can compete in the digital era (Unctad 2018).

For the dominant online platforms are mainly headquartered in China and the U.S., at different stages of development but both with abundant high-tech talent, a large domestic market, and frontier capabilities in the ICT sector. However, the two countries have different regulations on data and cross-border data flows, which can also affect the relative comparative advantage of their digital firms. Traditional trade theory indicates that most high value-added activities, such as R&D, are mainly located in developed countries with firms covering the fixed cost of these activities by selling the products with their embodied R&D outputs to developing (as well as developed) countries (Klepper, 1996). That is, most new product innovations do not rely on the local inputs in developing countries.[2] However, data is a key input for firms to develop and deliver digital services and products to developing countries, whose consumers are the data source. This is different from earlier types of innovation, where the consumers in developing countries did not in general contribute to the innovation process.[3]

Moreover, the technical barriers to entering digital markets can in some cases be lower than barriers to entry in some traditional products requiring significant financing, highly skilled labor and sophisticated supporting supply chains. So the degree of technology sophistication required by local entrepreneurs can be low and the required minimum operational scale can be small, as long as they can get access to cheap cloud computing services. Some skills and finance will be required but the key is to have a viable data-driven business model, with data access, serving the local market. Entrepreneurs can get access to more affordable and adaptable advanced AI tools without the need to have their own advanced know-how, as long as they have the data to feed the algorithms. Previously, physical capital-intensive industries (such as the auto industry and the information technology hardware industry) have required a large domestic market or export markets to be able to amortize the heavy capital investment costs. Now, subject to regulatory barriers, even start-ups in developing countries can tap overseas markets relatively easily. China's TikTok is an excellent example both of the possibility of expansion in overseas markets, and of the impact of trade and investment policies. In addition, small and medium sized firms (SMEs) can tap overseas markets relatively easily by utilizing the services provided by e-commerce platforms. Lastly, local entrepreneurs may have a comparative advantage in understanding local unserved demand and/or the problems faced by local consumers and businesses, developing indigenous business innovations to solve the local needs better by using the insights derived from big data. There are plenty of examples in developing countries such as Gojek, Kudo and Tokopedia in Indonesia. Thus Kudo was founded by a local entrepreneur who identified a cheap way to connect over 400,000 small mom-and-pop grocery stores around the country which has more than 17,500 islands, turning the stores into online shopping and distribution stations to address the challenging logistics and credit problems for local people.

As noted, the data advantage of dominant online platforms can act as a formidable entry barrier in developed and developing countries alike; but these are not insuperable. As shown by the case of Airbnb challenging Booking.com, new startups may not be deterred by existing dominant online platforms as long as they can find a niche through which to enter. In addition, for most digital start-ups their buyout by bigger platforms is an exit option, while dominant online platforms regard startups as a source of innovations (Coyle et al., 2020). Dominant online platforms can serve as a bridge between local start-ups, small businesses and international markets.

Thus most high value-added activities can now be carried out in one or a few locations while digital products and services produced by AI algorithms can be delivered worldwide at near-zero transportation cost. Does this mean that Adam Smith's principle that the division of labor is limited by the extent of the market is no longer valid in the winner-take-all digital era? That is, as the digital markets grow, will we see increasing geographic concentration of most value-added activities rather than the international division of labor and increasingly segmented supply chains in physical production observed during the era of globalization since the 1980s (Li et al., 2019; Morton et al., 2019)?

There are drivers in both directions. Digital industries may become more concentrated, with the U.S. and China the world's digital 'factories'. Countries with a bigger population will definitely enjoy a comparative advantage in data. However, as argued above, there are limits due to national market differences. For example, the patterns recognized in Chinese consumer data will not necessarily apply to services for consumers in the U.S. (Dvorak, 2018). Given that data is the key for the producing and developing digital products and services in a country, even for dominant online platforms, accessing local data is essential.

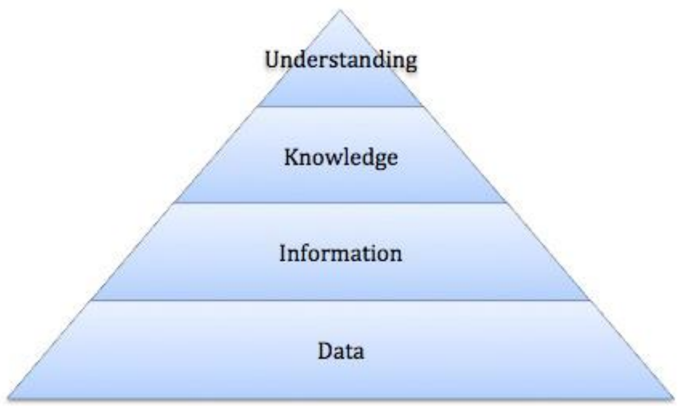
The existing literature thus raises important issues. How should we conceptualize data value? How can we measure the size and growth of data markets? And what framework will help analyse data trade flows? We begin with data value, describing how it is formed, and setting out a methodology applied here to the example of the hospitality sector. The method can be applied to other sectors or to countries. We then propose a typology of countries as a framework for thinking about data trade flows, with case studies for each type.

3. Understanding and Measuring the Value of Data

3.1 The Value of Data and Sharing Data

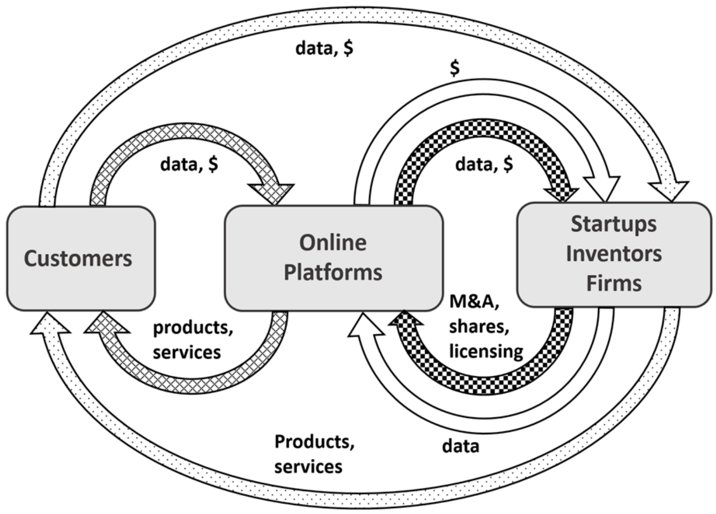
The previous section set out the potential for data to form a key part of organizational capital in the digital economy. Here we clarify what we mean by the value of data, and how it is created. As set out in the classic information pyramid (Figure 2), the value of data depends on processing it to enable firms to make better decisions, enhancing their competitiveness, growth and profits. The value of data lies in the firm-specific knowledge derived from the use of data. Raw data itself, and even structured databases as currently included in some national accounts measures of intangibles, will not have much value but is nevertheless now a key ingredient in innovation and production, including the innovation of digital goods and services, and digital fabrication. Firms that collect data have legal ownership but it is their utilization of data rather than the ownership that creates economic value (Statistics Canada 2019). We are measuring the value of data in use, as the economically meaningful concept driving productivity and comparative advantage.

Figure 2: The Information Pyramid



As there is a virtuous cycle between data and AI algorithms enabling better digital products and/or services, data involves significant positive externalities (Coyle et al 2020). Online platforms have internalized the positive externalities (Li et al., 2019). Even they will, though, face the limitations that arise from the restricted types of data they collect through their own platform. Thus there is considerable potential productivity gain in increasing data sharing for all market participants (Figure 3). The positive externalities create the incentive for data sharing and data trades.

Figure 3: The Virtuous Circle of Data Sharing



Source: authors' own

The bigger the pool of data, the more participants, more innovations, and greater positive externalities exist. New economic value can be created when new combinations of data enable entrepreneurs, inventors, and firms to derive new insights. In addition, data sharing will enable a greater degree of entry and competition, driving more innovations. The positive externalities associated with data sharing help increase the social returns to data at an increasing rate (Jones and Summers 2020). An open data-sharing ecosystem with an appropriate framework for access rights can unlock innovations in how we produce and deliver the products and services, eventually resulting in the increase in productivity and economic wellbeing.

Even dominant online platforms with hoards of data have some incentives to trade and/or share data. There are several reasons. First, online platforms also need different types of data to enter new markets to expand and/or diversify their businesses. They otherwise are limited to the types of data that they can collect on their platforms. For example, in contrast to Amazon, Facebook initially did not collect actual transaction data on its users' purchases, and therefore has aimed to cooperate with banks to access consumers' financial data to better understand user behavior. Facebook eventually created the Libra project, a digital currency project collaborating with major financial incumbents to create a new ecosystem to enter various businesses beyond its current major data-targeted advertising business (Murphy, 2019). Another example is Amazon's request to access the data of its invested start-ups and third-party sellers on its platform in order to develop its own new products and services (Mattioli and Lombardo, 2020; Mattioli, 2020).

Second, because start-ups are a major source of new innovations, dominant online platforms have become a venture capital funding source, such as Amazon’s Alex Fund and Alphabet’s GV (Sullivan, 2018; Kolodny, 2017). Some may allow their invested start-ups access to their data to test new AI algorithms and new data-driven business models, sharing the data under an appropriate mechanism. More tests on invested start-ups’ AI algorithms and data-driven business models mean a higher chance that one of the ventures will succeed, and therefore higher investment returns. Start-ups may also collect novel data that dominant platforms would find useful.

At the aggregate level, the more firms share data, the more innovation can be expected within and across firms, industries, supply chains, sectors, and countries. This is increasingly true as more and more products will be equipped with sensors that collect a variety of data in real time. All market participants across industries will benefit from access to data. For example, manufacturing operations can be managed remotely if manufacturers can access cross-border data in real time. Another example is health data which could be collected across borders and help develop drugs for under-served groups or countries. Data-driven ecosystems thus have the potential to create significant economic value.

These considerations concerning the source of the value of data including externalities and the potential for innovation and growth are rarely taken into account in national data policies, however. Countries have very different data sharing regulations. At one extreme is China where “special management shares” are used by Chinese government to manage digital firms; the government can invest in a 1% stake to get a direct role in the strategic direction of the company (Yuan, 2017). In addition, although China has several active public data markets, firms need to get government approval to trade. More importantly, the Chinese government has established private networks to allow selected strategic businesses, including all Chinese big tech companies, to exchange and share data without restrictions, a mechanism that gives Chinese businesses a great comparative and competitive advantage in innovations, including in business models and AI (Initium Media, 2019).

At the other extreme, the EU’s GDPR restricts firms from repurposing data beyond its original intended use without re-obtaining consent from individuals, to safeguard privacy, which limits data sharing among firms and countries within the EU. In addition, whether because they want to keep a dominant market position or because they do not understand the value of their data, many western firms are reluctant to share data with other firms. Most data is collected for own use. However, the EU's GDPR does allow firms to transfer data to another country that has a similar data privacy protection measures, whereas China does not allow data transfer across borders without official review and approval. In practice, it is difficult to transfer data from China to other countries. Data sharing can thus be open within a country but restricted across borders or the other way around. Restrictions on data transfer for reasons of trade protectionism seem to be on the increase. The implied transaction costs involved for firms, especially for SMEs, to conduct digital trade can be tremendous. Therefore, it is very important that we can measure the size of the market taking into account that the value is use value, rather than inherent in the raw data, so that businesses and policymakers can have a sense of the magnitude of the market size when they make decisions.

3.2 Measurement Methodology

No accepted methodology to measure the value of the market for data exists currently. Apart from the significant conceptual challenges (Coyle & Diepeveen 2020, Nguyen & Paczos 2020), the biggest hurdle is the dearth of market prices from exchange; most data are collected for firms' own use. In addition, firms are unwilling to release information relating to transactions in data, such as the private exchange or the sharing of data that occurs among China's big tech companies. There is no data on detailed cross-border data flows, either at the industry level or at the firm level. Moreover, to be able to estimate the value of data, we need to estimate its depreciation.

Here, we propose an "impact-based" approach to estimate the size of data market by comparing the values of data with and without the entry of an online platform in the industry of interest. The value of data for a firm is estimated by its organizational capital, which as mentioned earlier represents the accumulated information of the firm. As shown below, the entry of an online platform can accelerate the depreciation of the organizational capital of a traditional firm without digital transformation, resulting in a reduced organizational capital as compared to the scenario where no online platform entered the industry sector. This differs from an alternative approach based on the change in the market capitalization of firms identified as ‘data driven’ (eg Ker & Mazzini 2020) in that we identify and estimate specifically the organizational capital component. While our approach may also include the impact of other changes in organizational capital, such as improved management quality, it is less likely to be affected by factors other than data and data use as compared to the entirety of firms’ market capitalization. Our new approach makes use of the recent finding that the entry of an online platform can create a disruption in the organization capital of the incumbents by affecting its depreciation rate (Li and Chi, 2021). As the value of data lies in a firm’s specific knowledge derived from that data, it can be measured by the organizational capital of the firm, or in other words the accumulated information that guides its decisions concerning how to produce, compete, and grow (Li and Chi, 2021; Prescott and Visscher, 1980). Li and Chi (2021) studies the impact of the entry of online platforms on the incumbents in the US hospitality and transportation industries. They find that when online platforms entered those industries, the disruption mainly resulted from incumbents’ relatively limited amount of data, and consequently their limited useful firm-specific knowledge, compared to the new entrants. The depreciation of their organizational capital is therefore a measure of the impact of disruption.[4]

The loss of the value of incumbent firms’ organizational capital due to their data disadvantage can be used to measure the potential size of the demand for data by such firms in the industry sectors disrupted by online platforms. That is, specifically, we measure how much firms should be willing to pay in order to maintain the value of their firm-specific knowledge derived from data. We can thus use the loss of the value of organizational capital to measure firms’ maximum willingness to pay for the access to data.[5]

We proceed as follows. We apply the Li and Hall (2020)[6] depreciation model to first estimate the depreciation rates of incumbent firms’ organizational capital before the entry of an online platform; we then estimate the after-entry depreciation rates. The Li and Hall (2020) model is a forward-looking model that only requires firm-level data on sales and investments in intangible capital to identify the firm-level depreciation rates of such intangible capital. Then, based on Li and Chi (2021), we assume that the depreciation rates of organizational capital by incumbent firms can be maintained at prior rates if they undertake their own digital transformation. This assumption is consistent with the principle of intangible capitals that, in the same industry where firms face the same competitive environment, the main cause of the depreciation is obsolescence.[7] Therefore, we can adopt Hall’s (1993) methodology to calculate the stocks of organizational capital based on before-entry and after-entry depreciation rates. We use the difference between the two stocks as the proxy for the demand for data by disrupted firms. This difference measures the loss to these firms due to their failure to use data in order to cope with changes in competition due to the entry of an online platform.

3.3 Our Data

In this paper, we use the firm-level data for the hospitality industry between 2002 and 2019 from the Compustat database. We study the impact of the entry of Airbnb on existing firms in the industry. Airbnb has enabled new supply of short-stay, private accommodation, and while the hospitality market had expanded (pre-pandemic), it was increasingly competing directly with hotels for business stays. Studies indicate it has led to a decline in bookings and revenue losses for hotels in some markets (Blal et al 2018, Farronato & Fradkin 2018). The hospitality industry is one of the leading service industries affected by online platforms. Some incumbents such as Marriott have a platform business model but not one based on big data. Looking at these incumbents can demonstrate the scalability of online platforms compared traditional players adopting similar business models but not based on big data. The industry is also of interest because it is one of the traditional industries in which most companies do not report investment in R&D yet have been deeply disrupted by online platforms. Finally, the travel sector has also been severely disrupted by the COVID-19 pandemic, due to the drastic change in consumer behavior, so the lessons learned from its urgent need for digital transformation may be useful elsewhere.

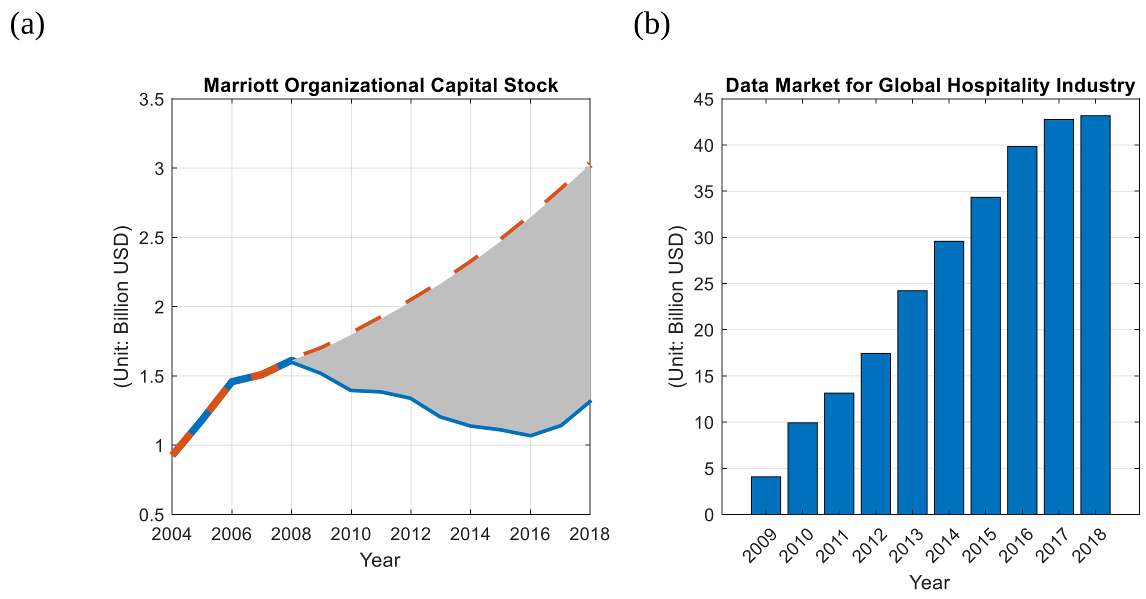
3.4 Empirical Results

Our data analysis starts by estimating the depreciation rates of Marriott's organizational capital over time using the Li and Hall (2020) model. Because Airbnb entered the hotel industry sector in 2008, we used two segments in Marriott's organizational capital time series, one from 2003 to 2008 and the other from 2010 to 2015, to calculate the depreciation rates before and after Airbnb's entry. We found that the depreciation rate of Marriott's organizational capital rose from 45.2% in the "before" phase to 52.9% in the "after" phase, a result consistent with the finding in Li and Chi (2021) that the disruption by a new online platform would lead to faster depreciation of the organization capital of incumbent firms. In addition, Li and Chi (2021) have recently shown that, in contrast to the case of Marriott, online platforms such as Booking and TripAdvisor experienced slightly reduced organizational capital depreciation rates after the entry of Airbnb. Considering the results from these firms in the same industry sector and the same competitive environment, we conclude that the main cause of the change in Marriott's organizational capital depreciation rate is the threat to the firm's business model.

We then used these depreciation rates to consider two scenarios for Marriott's organization capital stock time series: one based on the actual depreciation rate that declined after Airbnb's entry, and the other based only on the depreciation rate for the "before" phase. The latter hypothesizes a case as if Airbnb never entered the market. Therefore, the two organizational capital time series start to differ after the year of Airbnb's entry, and the difference, as shown by the shaded area in Figure 4a, can be taken as a measure of the potential demand for data. We note that the rise in the organizational capital stock in 2017-2018 is associated with the increase in relevant investments after Marriott grew by acquiring Starwood.

To estimate the potential value of the data market for the global hospitality industry, we considered the proportion of Marriott's market share. Specifically, we divided the shaded area in Figure 4a by Marriott's share of the approximately USD $600 billion (2018) global market.[8] The result shows that the market size for data in the global hospitality sector is substantial, reaching USD $43 billion in 2018 (Figure 4b). This data market has also grown rapidly at an average growth rate of 35%, meaning that its size has been doubling in less than three years. Because Marriott is the leader in this sector and thus expected to have a lower organizational capital depreciation rate than those of its followers (Li, 2015), we note that this market size is an order-of-magnitude and likely conservative estimate.

Figure 4: Marriott's organization capital stock and the estimated data market size of the global hospitality industry



Source: authors' own

Although this paper considers only one industry as an exemplar of the method, our approach provides a way to measure the value of other data markets and has several important implications. First, it provides a magnitude for us to start to understand how big the market for data for an industry can be and how quickly it can grow, both having important policy implications for thresholds selected for data localization requirements. Second, the rapid market growth implies that online platforms' disruption of incumbents' firm-specific knowledge is fast and significant, especially given that the accumulation of data can have increasing returns. Third, a growing number of online platforms are adopting China's super-app model (offering a range of services through one app), meaning they can collect various types of data with greater scope for combining them, with resulting advantage in business innovation and data-targeting services (or mass customization). This implies that their impact may be accelerated or multiplicative. Hence, in addition to the possible competition implications, the market for data may grow even faster when we consider markets across industries and across countries.

4. The International Division of Labor: A typology of data trade

In this section, we introduce a typology of trade in data, and digital goods and services. This will help explain trade patterns, and the necessary investments and infrastructure, in the data economy and can inform the design of policies for data trade.

Before the IT-enabled outsourcing era, multinational firms captured the benefits of globalization by exploiting increased global market opportunities, thus spreading their R&D and other costs over international markets, either producing in-house and/or outsourcing production overseas. Offshoring production of physical goods helped raise the level of technological performance in developing countries; the technologies, however, are still tied to advanced countries. In the IT-enabled outsourcing era, offshoring in high-tech industries provided a new market mechanism for contracting firms in East Asian countries such as Taiwan and China to invest in R&D and generate indigenous technological innovations. This has meant that not only are downstream products tradable but also upstream R&D jobs, making the traditional North-South assumption in Vernon's product life cycle theory (Vernon, 1966) outdated. Li (2008) demonstrated the existence of a virtuous cycle between offshoring demand and the contracting firms' technological expertise. This cycle helps explain why U.S. IT firms have been increasingly outsourcing innovation overseas, especially to China and India. Since early 2000s, western high-tech firms have been setting up research labs in these two countries. Moreover, Li (2008) also points out that if we combine the facts that (1) offshore outsourcing contracts have been flocking to regions like China and (2) those countries' skilled expatriate scientists and engineers are increasingly returning after study or work overseas, this combination has provided opportunities for firms in China to move up the value chain and develop new industries, especially where standards and markets are not yet established, such as 5G equipment and AI. Both China and India not only have abundant tech talent but also a large domestic market to incubate 'national champions' in these new areas.

Both R&D and data are intangibles but developing countries play different roles in contributing to the creation of the value of each type of intangible capital. Most R&D assets are created in developed countries and concentrated in rich countries, with a few exceptions such as China and India. However, the key ingredient, data, must come from the consumers and in future the IoT sensors located in developing countries themselves.

Here, we classify six basic types of trade in data, and in digital goods and services, looking at countries by the market size, their degree of economic development, the readiness of their digital infrastructure, and the presence of dominant international online platforms. By data we mean the underlying data records, which have been a focus in policy discussions (eg Unctad 2021).

We identify six basic types of trade in data, and digital goods and services, based on the following distinctions: whether a country is a *net (raw) data importer with existing dominant global platform companies or a data exporter*; whether it is a *developed or developing economy (corresponding to the World Bank's definition of high income countries and others)*; whether it has a *large or small domestic market (dependent on population and income per capita)*; whether it has *other high-tech advantages* including talent and digital infrastructure. Assignment of any specific country to one of these requires an element of judgment.

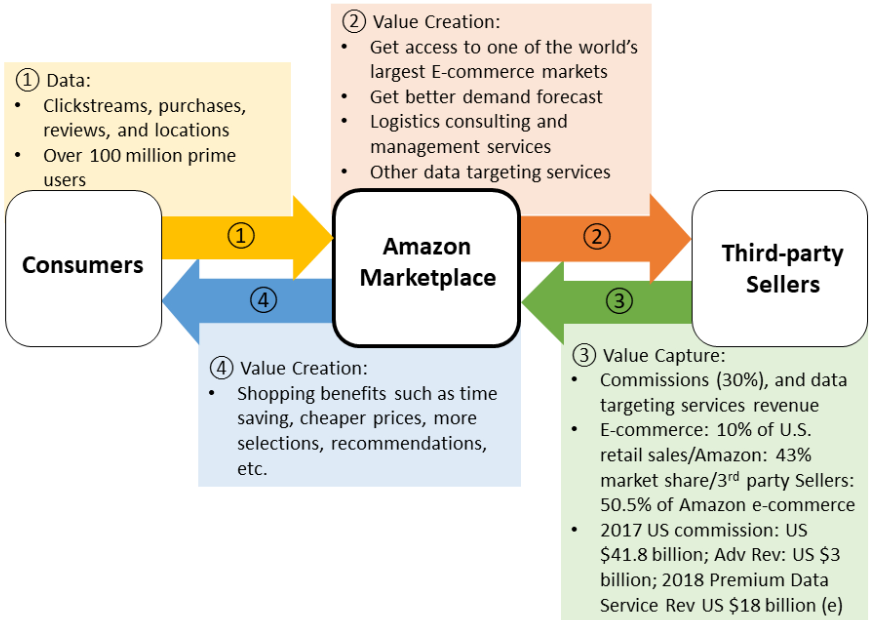
Type I: Net Data Importers – Large Developed Countries with Dominant International Online Platforms and Leading High-tech Industries
Type II: Net Data Importers – Large Developing Countries with Dominant International Online Platforms and Leading High-tech Industries
Type III: Net Data Exporters – Large Developing Countries without Dominant International Platforms but with Leading High-tech Industries
Type IV: Net Data Exporters – Large Developing Countries without Dominant International Online Platforms and Leading High-tech Industries
Type V: Net Data Exporters – Developed Countries without Dominant International Online Platforms but with Leading High-tech Industries and/or Talent
Type VI: Net Data Exporters – Small Developing Countries without Dominant International Online Platforms or High-tech Talent

Type I: Net Data Importers – Large Developed Countries with Dominant International Online Platforms and Leading High-tech Industries

Type I consists of one large developed country – the U.S. The U.S. has global platforms that collect data around the world and centralize decision making and digital production in the US. Therefore, the U.S. is a net data importer and a net exporter of digital goods and services. We use Amazon to illustrate how a dominant international online platform, a data-driven multinational firm, utilizes data to derive economic value, what role data plays in international competitiveness and innovation, and the resulting trade pattern of data flow and digital goods and services. Li et al. (2019) examine Amazon's data-driven business model from the perspectives of data flow, value creation for consumers, value creation for third parties, and monetization by Amazon (Figure 5). Amazon collects data from consumers and conducts data analytics to provide services to third-parties. It also uses the insights derived from data to create new products and services for its users, including personalized subscription services and recommendations for all products and services sold on its marketplace. For example, in 2019, Amazon had more than 100 million Prime subscribers in the U.S., a half of U.S. households, with the revenue from the annual membership fees estimated at over US $9 billion (Tzuo and Weisert, 2018). As shown in Figure 5, Amazon is able to take advantage of the feedback loops its business model creates by creating additional value beyond that generated by the chain of transactions and thus effectively captures much of the social value of the data they have accumulated.

Amazon is currently the largest e-commerce platform in the west but also serves some Asian countries. It had a 13.3% of worldwide retail ecommerce sales in 2018; Germany is its largest foreign market (Enberg, 2018). Most of its offices in other countries are warehouses, data centers and sales offices. In terms of data flow, Amazon collects raw data from its overseas customers and provides value-added digital goods and services to overseas consumers and third-party sellers on its platform. It has one of the most sophisticated logistics systems worldwide to support its global e-commerce activities, a key advantage. Amazon's AWS is one of the world's two leading global cloud service platforms with a highly sophisticated multinational data infrastructure. It has more than 100 data centers spread across 15 cities in 9 countries in 2015. 61 in the US, 14 in Europe, 35 in Asia-Pacific region, and 6 in Brazil (Wiki, 2015).[9]

Figure 5: Type I: Amazon



Source: Li et al. (2019)

Amazon thus has the key ingredients for success in the digital era. For example, its advertising revenue has been growing rapidly in recent years because its troves of data include not only browsing data but actual transaction and logistics data, in contrast to Google's search data. Amazon collects data through its online websites, its physical stores such as Whole Foods and Amazon Go, and stores it in its own AWS data centers. The data-driven business model is supported by its supply chain and logistics system (Fox, 2020). Amazon has numerous warehouses located in many countries while centralizing the decision-making in its US headquarters. Its large group of in-house data scientists, economists and engineers work in either in the U.S. or a few offices in advanced economies.[10] However, Amazon needs to collect data from developing countries to operate its global online businesses in those territories.

National data governance rules affect its global operations. For example, China has the largest e-commerce market among developing countries, US $1.935 trillion in 2019, almost 53% of U.S. e-commerce market size the same year.[11] Amazon accounts for less than 1% of China's retail ecommerce market while Alibaba had 58.2% in 2018 (Enberg, 2018). China's newly-enacted data laws since 2016 require firms that move data overseas to provide legitimate reasons and get local government approval in advance. In practice, any data related to China's citizens needs to be retained in China. In addition, foreign firms in general are not permitted to apply for data center licenses, which led AWS to sell its cloud servers to its local partners, Beijing Sinnet, for US $300 million in 2017 (Lin, 2018; Foerster, 2020).

Amazon, as a global online platform, enjoys positive externalities generated from the virtuous cycle between AI and big data and the network effects of its online platform (to the extent that it faces anti-trust actions in some jurisdictions including the EU). For example, Amazon Marketplace provides users more convenient choices at cheaper prices; it gives sellers a bigger market in a cost-effective and time-efficient manner. Amazon also provides data-targeting services to sellers to improve areas such as sales forecasts and logistics management. When Amazon enters markets, it can provide startups and firms of all sizes an opportunity to access to a large global market with low transaction costs. It also provides them with some data-targeting services (built-in data analytics services), and cloud computing services, a combination that lowers entry costs for new firms. Additionally, dominant platforms like Amazon with huge cash piles have become major players in providing venture funds to startups in developing countries, as an effective way for them to develop new innovations and/or enter new markets.

Type II: Net Data Importers – Large Developing Countries with Dominant International Online Platforms and Leading High-tech Industries

Type II consists of large developing country net data importers, the one current example being China. Our case study is TikTok, short video-sharing app launched by ByteDance in 2017 to target markets outside China. It has become one of the most popular social platforms in the world and the largest unlisted tech unicorn, valued at $90-$100 billion in early 2020 (The Economist, 2020). It has 800 million active users worldwide and ranked as the most-downloaded app on the Apple store in 2019. In addition, it is the only tech company bar Apple with more than 100 million users both in China and in the West (The Economist, 2020). TikTok is available all around the world via the Apple App Store or Google Play store. It has been growing especially rapidly in the U.S. and India. In the U.S., the number of adult users was doubling every 7.3 months until 2020, while 25% of downloads come from India. To date, it has operated in more than 150 countries and regions and is available in 75 languages (Fannin, 2019). TikTok's investors include venture capital firm, Sequoia Capital China, U.S. private equity firms, KKR, Japan's Softbank

Group, Chinese investment firm, Hillhouse Capital, and corporate venture unit SIG Asia (Fannin, 2019). Abundant funding enabled TikTok to spend almost US $1 billion a year on Facebook advertising to attract users (Wells and Kubota, 2019).

On one side of the market, it provides individuals with way to get free entertainment and news instantly. On the other, it allows advertisers and content providers to reach conveniently one of the world's largest user bases. It also enables content providers to monetize their content through advertising, partnerships with firms, and deals with talent agencies. TikTok collects user data on search terms, revealed preferences, browsing behaviors, locations, demographics, languages, and more. It conducts data analytics to provide data targeting services, such as targeted advertising, for third parties, and customized video feeds to its users. In addition, it also sells digital goods such as emojis and stickers to users.

Douyin is the Chinese version of TikTok, launched earlier in September 2016. Both use the same software but maintain separate networks to avoid Chinese censorship restrictions (Moshin, 2020). Although ByteDance keeps two separate networks, TikTok transfers data back to ByteDance's servers (Pousen and McMillan, 2020), and the combined number of active users in both U.S. and China give TikTok a great advantage in data, improving its algorithms. For example, in 2020, its U.S. social media rival Facebook has an estimated 223.03 million active users in the U.S. and 3 million (via VPNs) in China. TikTok and Duoyin have 80 million active users in the U.S. and 400 million in China respectively, a bigger combined total. Compared with competitors like YouTube or Instagram, TikTok's algorithm is better at quickly figuring out what types of music and videos get users' attention (Stern, 2020). Moreover, the open data sharing environment among Chinese enterprises also gives ByteDance a better environment to test and develop smarter algorithms than western platforms. TikTok can also collect data from India's 200 million users and users in other countries, whereas its foreign counterparts are banned in China. ByteDance's main western competitors are all banned in China, including Facebook, Twitter, and YouTube (Yang, 2020). TikTok entered the U.S. market without restriction, although the regulatory environment has clearly changed adversely since then. France is investigating whether TikTok violates the EU's GDPR in its use of personal data which could lead to additional restrictions on trade in data and digital services with China. The climate for cross-border data sharing is changing.

Type III: Net Data Exporters – Large Developing Countries without Dominant International Platforms but with Abundant High-tech Talent

Unlike China, India does not have strict restrictions on cross-border data flows and its market is open to Western platforms.[12] As a result, Amazon and Walmart are leading players in the Indian ecommerce sector. Facebook's WhatApp is one of India's main communications tools. India is Facebook's largest market in terms of the number of users. In digital payments, Google is a leading player in India's market. However, Western and local competitors still face a competitive data disadvantage vis à vis Chinese platforms.

For instance, Jio is a local online platform and the digital service subsidiary of India's largest network operator, Reliance Industries. Jio has adopted a super app strategy to compete with U.S. and Chinese online platforms; however, all of its funding sources are foreign, including U.S. platforms such as Facebook and Google (Sender, 2020). Although Jio Platforms has significant investments and 400 million users in India, its user experience lags behind Chinese and U.S. competitors. Interestingly, Amazon also has a super app strategy in India and competes fiercely with Jio in groceries, pharmacy, entertainment, food delivery, insurance, and wealth management (Lin, 2020). Jio Platforms is a local major player in e-commerce and other sectors but it cannot compete with international dominant online platforms outside the Indian market.

In terms of funding, compared with their Chinese counterparts, Indian firms rely more on foreign investors. China's Alibaba and Tencent have become the key sources of venture capital not only for startups in China but also in India, Southeast Asian countries, and the West too. Since early 2017, both the Chinese companies have invested aggressively in India. There are many Chinese apps growing in popularity in India, among them TikTok, which has over 200 million users there and 2000 local employees (Business Weekly, 2020).

The current dominant online platforms in key sectors are Chinese and Western platforms. Therefore, India is a net data exporter but foreign direct investments from Chinese and Western firms enable local firms to enter the market. Moreover, those foreign investors may also provide opportunities for strategic alliances to access to foreign markets and technology or business knowhow.

Type IV: Net Data Exporters – Large Developing Countries without Dominant International Online Platforms or Abundant High-tech Talent

For large data-exporting developing countries without abundant high-tech talent, like Indonesia, there is nevertheless a chance that they can grow their own successful domestic online platform companies. Indonesia has more than 17,500 islands of which about 6,000 are inhabited. It consists of 5 major islands and about 30 smaller groups, a geographical feature that poses special challenges for logistics networks for e-commerce. Moreover, 60% of population did not have bank accounts in 2017. These conditions posed a big challenge for building ecommerce services. However, Indonesia has developed its ecommerce by investing heavily in basic infrastructure,[13] lifting restrictions on foreign direct investment to allow foreign investors a 100% share, and attracting big global online platforms to help the buildup of ecommerce businesses. It is now the largest ecommerce market in ASEAN with a market size estimated at US $65 billion by 2022 (Widowati, 2018). Around 30% of the 7,000 plus start-ups in ASEAN are located in Indonesia (Guan, 2017).

Despite aggressively investing in Southeast Asia, Amazon does not have a major role in the Indonesia ecommerce market while Alibaba has not been able to defeat the local unicorn, Tokopedia. Tokopedia, founded by local entrepreneurs in 2010, is the largest ecommerce marketplace in the region (ASEAN UP, 2019). It uses the super app business model and has operations in areas including fintech, digital payments, and logistics. Its majority investors are international VCs including US Sequoia Capital and SoftBank. In 2017 and 2018, Alibaba and SoftBank increased their investments in the firm (Tokopedia Wikipedia, 2020). In addition, international cloud service providers such as Alibaba and Amazon have aggressively invested in the region, allowing local entrepreneurs access to cheap cloud computing resources (Chan, 2018). But like Indian platforms such as Jio, Tokopedia would face daunting competition from Alibaba and Amazon, with their huge advantages of cash, data and know how, if it tried to enter international markets.

Facebook and PayPal have also recently invested in Gojek, a leading Indonesian on-demand multiservice platform and digital payment group (Wang, 2020). Big global platforms can therefore help stimulate innovation and the creation of new businesses. A key advantage of local entrepreneurs in developing countries is their deeper understanding of local unmet demand. With a supporting digital infrastructure, SMEs and other firms in developing countries can outsource data storage and analytics to international cloud computing service firms like Amazon or Alibaba.

The technical entry barrier is much lower than in the past in countries like Indonesia, without abundant high-tech talent.

Type V: Net Data Exporters – Developed Countries without Dominant International Online Platforms but with Leading High-tech Industries

Countries like France and UK have no dominant local platforms. Countries in this category are net data exporters. They have good digital infrastructure and good data analytics capabilities, but a disadvantage in data. Few countries have large global platforms. According to Weber (2017), based on Faravelon et al. (2016), only 11 countries at that time hosted influential online platforms. The U.S. had 32, China 5, and a few other countries one each. In terms of the distribution of influence, U.S. platforms have a nearly global reach, Chinese platforms are big players in a small number of countries (fewer than 10), some Brazilian platforms are big players in an even smaller number of countries, and the numbers go down from there. For example, 70% of online platforms visits in the UK and France are to overseas online platforms (Faravelon et al. 2016). Weber points out that there is a real risk of self-reinforcing dependency that traps countries like France, although a developed country, in a data periphery role, just as some countries have previously been trapped as a low value-added raw material exporter and high value-added data product importer. The question is what options these countries have to develop long term economic growth in the data economy. Weber was pessimistic about the possibility for data peripheral countries to leapfrog or catch up with China and the U.S.

Since 2017 the picture has changed somewhat. Some of Chinese popular apps have significant market shares in the west or even outperform their Western counterparts in other non-China markets. Platforms such as Alibaba and Tencent have become major sources of venture capital not only in China but also in foreign markets including in the West. Moreover, in contrast to the superapp business model, it is easier for local entrepreneurs to copy the business model of a simpler online platform model like ridesharing, and to gain first mover advantage thanks to local regulations and market conditions. Several ridehailing apps in Southeast Asian countries outperform Uber and Lyft, as does China's Didi. So it is possible for local online platforms to compete with international dominant online platforms in their home markets. But it may still not possible for them to compete in global markets for the reasons set out above. Just like firms in

Indonesia, local firms in the UK and France can have a disadvantage in data and finance but an advantage in creating business models serving their domestic market due to their deeper understanding of the bottlenecks facing existing businesses and unmet consumer demand. Small startups may not be able to compete with dominant international online platforms in international markets but the funding sources from international dominant online platforms or other big tech companies can help incubate the new ventures.

Type VI: Net Data Exporters – Small Developing Countries without Dominant International Online Platforms or High-tech Talent

For smaller developing countries such as Vietnam, there is no restriction on the entry of international platforms. The top five websites in all categories are Google, Facebook, YouTube, he069.net, and Google.com.vn (SimilarWeb, 2020). Facebook in Vietnam has over 20 million more users than it does in UK (Yang, 2019). Therefore smaller developing countries like this are net data exporters but their market size can be larger than some developed countries for the big global platforms. They rely additionally on foreign direct investment to invest in startups and basic digital infrastructure.

Some small developing country markets such as those in Africa are net data exporters but are far less well placed. They are in the data periphery category. Most African countries have limited basic digital infrastructure and little or no access to cheap cloud computing services. Small developing countries in this situation face prohibitive costs of building and managing data centers and infrastructure (Munshi, 2020). Africa's current data centers only account for less than 1% of global capacity even though it has around 17% of the global population, a gigantic potential data generator. If small African economies were to adopt China's data center restrictions – and data localization is starting to be implemented – foreign cloud service companies and online platforms might not have enough incentive to invest and enter the market. Local firms would be even less likely to be able to access to tap the potential of data and AI.

Even in South Africa, the highest income per capita country in Sub-Saharan Africa, Amazon currently does not offer e-commerce and only launched its first data center operations there in April 2020, an operation that relies on independent firms' data centers, e.g., Africa Data Centers of the South Africa-based subsidiary of Liquid Telecom (Reuters, 2020; Munshi, 2020).

4.3 Discussion

The key upstream input for delivering valuable digital goods and services is data. We have established through the example of one sector, hospitality, that the value of data is large and growing rapidly. A distinctive aspect of data compared to other intangibles such as R&D is that firms alone produce R&D whereas access to consumers or users is required to provide data and generate value, even though legally it is generally owned by the firm collecting it (Jones and Tonetti, 2020). The value created by using data to produce digital goods and services depends also on skills and financing, concentrated in a few countries. But all countries, even developing countries, will be essential data providers for businesses serving their markets.

This feature has implications for trade in data and for potential market entry, innovation and growth. Both upstream inputs and downstream products are tradable, where there is no restriction on data transfer across borders or between firms. Moreover, both inputs (data) and outputs (digital products and services using data) can be delivered at almost zero (marginal) cost and in real time. This implies that digital goods and services can be produced anywhere as long as there is access to the data. The traditional comparative advantage of developing countries of lower labor costs may be weaker. Unlike a physical supply chain, local clusters of related suppliers are not necessary for a data value chain, remote work is possible. When remote work is possible on a global scale, the required degree of international skilled labor mobility will be lower for the global data economy.

Developing countries are unlikely to be able to invest in costly digital infrastructure such as data centers and rely heavily on international investors to incubate and develop e-commerce and other platforms, as well as to build costly logistics and warehouse networks to support ecommerce. However, costly digital infrastructure can be accessed either by purchasing services from cloud providers overseas or through foreign direct investment by international companies. As a result, data imports and exports are not directly related to the creation or the distribution of the value of the data. Consider the role of data centers in the creation of the value of data. Because it can cost about US$1 billion to build a large data center, firms in some developing countries may have to rely on cloud services located in other countries, which will naturally increase the cross-border data flows. Yet the data flows alone do not constitute the value of data. For example, Google currently has two data centers in Taiwan to support its operation in Asia. One of them supports

Google's provision of over 4 million apps to Asian countries (Ho, 2020). This means that there are large cross-border data flows between Taiwan and other countries, but Taiwan is unlikely to receive most of the benefits. The physical data records are delocalized but the use is not. Moreover, because the creation of value from the use of data requiring local know-how is the sole activity that cannot be outsourced, in a large developing country like Indonesia without abundant high-tech talent, local entrepreneurs can establish successful ventures. These indigenous ecommerce platforms can attract overseas third-party sellers to sell goods on their platform. This part of business activity increases cross-border data flows and conversely having a data-driven business model taps the growth of the cross-border data flows (Li and Chi, 2021).

Once there are dominant platforms in a market, no matter whether they are domestic or international, all potential competitors face the data entry barrier. As all firms need to access data, regulations on how data can be shared within borders and be transferred across borders, and the breadth and the depth of the markets for data, will affect the international division of labor.

Markets or frameworks for data access could reduce duplicate investments in areas such as data collection and storage, which involve expensive infrastructure investments. The depth of each market will determine the size of threshold above which a firm should invest in collecting and managing data in-house. More importantly, the positive externalities mean the breadth of data access can determine whether society can truly unlock the power of data to stimulate innovation and growth.

Varying rules in jurisdictions around the world can increase the transaction costs in markets for data, a topic that is beyond the scope of this paper. However, an appropriate framework for access to data could motivate data gatekeepers to invest in market mechanisms that increase the utilization of data. This would attract more participants, both suppliers and buyers, to enter the markets and hence increase market size in a virtuous circle.

5. Conclusions

In this paper we have discussed the impact of digital disruption on the scale and growth of demand for and use of data to create value, and described the motivation for increasing data sharing. We have demonstrated a novel demand-side methodology for estimating the size of data markets, applied to one sector. We have set out a typology of countries as a framework for future work estimating the size of data trade, illustrated with case studies. There are important policy implications as countries consider their data regulation frameworks.

This is a pressing research agenda for several reasons. One is that the amount of data gathered and potentially available to be used to create value is increasing rapidly. The Covid-19 pandemic has prompted globally a surge in the use of online platforms and e-commerce. The impending era of 5G communications and IoT applications in manufacturing and other arenas such as agriculture, healthcare, urban management or autonomous vehicles, and the increasing use of algorithmic decision making in public services, will drive further the use and potential uses of data.

Meanwhile, many countries are revising their data policies and localization rules. There is an emerging 'arms race' in AI, seen as a strategically important technology by the US, China and a number of other countries. The trend is for increasingly tight restrictions on data transfer. However, these policies should be informed by the implications of digitalization and data trade for international comparative advantage. First, the measurement of market size for data can help policymakers understand how big it is and how fast it can grow, and could inform thresholds for data localization requirements. Second, data imports and exports are not directly related to the distribution of the value of data or the creation of the value of data. Data flows alone do not reflect the value derived from data. Firms do not need specialist knowledge of advanced AI tools to build firm-specific organizational capital using data, in a high-value data-driven business model. Third, unlike a physical supply chain, local clusters of related suppliers are not necessary for a data value chain. Both upstream inputs and downstream products are tradable, where there is no restriction on data transfer across borders or between firms. Moreover, both can be delivered at almost zero cost and in real time, so digital goods and services can be produced anywhere as long as there is access to the data. Fourth, the value of data is created through its utilization, not its ownership per se. Also, since new values of data can be created through data fusion and through data-driven business innovations, data need not depreciate as long as one finds ways to use it, but the value of data created by firms may depreciate due to obsolescence and competition. An open data-sharing ecosystem with an appropriate framework for access rights could unlock innovation, eventually resulting in the increase in productivity and economic wellbeing.

A richer understanding of the location of value-creation and distribution from the use of data, combined with estimates of the size and growth of relevant data markets, would be helpful to inform the analysis of trade and the development of future data trade policies.

References

Arora, A., Fosturi, A., & Gambardella, A. 2001. Markets for Technology: The Economics of Innovation and Corporate Strategy, MIT Press.

ASEAN UP, 2019. Top 10 Ecommerce Sites in Indonesia 2019, https://aseanup.com/top-e-commerce-sites-indonesia/

Atkeson, A., and Patrick., J. 2005. Modeling and Measuring Organisational capital, Journal of Political Economy, 113, pp: 1026-1053.

Baldwin, R 2019. The Globotics Upheaval: Globalization, Robotics and the Future of Work. Oxford University Press.

Basul, A. 2019. Amazon Dominantes 30% of UK Ecommerce Market in 2019, https://www.uktech.news/need-to-know-2/amazon-dominates-30-of-uk-ecommerce-market-in-2019-20191213

Blal, Inès , Manisha Singal, Jonathan Templin, 2018. Airbnb's effect on hotel sales growth, *International Journal of Hospitality Management*, Volume 73, Pages 85-92,

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

Byrne, D., Corrado, C, and Sichel, D. 2020. The Rise of Cloud Computing: Minding your Ps, Qs and Ks,. in Measuring and Accounting for Innovation in the Twenty-First Century , eds Corrado, Haskel, Miranda, and Sichel.

Business Weekly. The New Cold War between China and India, and the U.S. Gains. August, 1709: 18-20.

Castro, D., & McQuinn, A. Cross-border Data Flows Enable Growth in All Industries, Information Technology & Innovation Foundation, February. http://www2.itif.org/2015-cross-border-data-flows.pdf

Chan, F. 2018. Amazon Planning $1.3 billion Investment in Indonesia, TheStraits Times, September 23rd.

Coyle, D., Diepeveen, S., Wdowin, J., Kay, L., & Tennison , J. 2020. The Value of Data: Policy

Implications, February,
https://www.bennettinstitute.cam.ac.uk/media/uploads/files/Value_of_data_Policy_Implications_Report_26_Feb_ok4noWn.pdf

Coyle, D., Diepeveen, S.. 2020, Creating and Governing Value from Data, mimeo.

Coyle, D & Nguyen, D, 2018. "Cloud Computing and National Accounting," Economic Statistics Centre of Excellence (ESCoE) Discussion Papers ESCoE DP-2018-19, Economic Statistics Centre of Excellence (ESCoE).

Demsetz, H. 1997. The Economics of the Business Firm: Seven Critical Commentaries, Publisher:
University of Cambridge.

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

Enberg, J. 2018. Amazon Around the World, eMarketer, November 13th.
https://www.emarketer.com/content/amazon-around-the-world

Fannin, R. 2019. The Strategy behind TikTok's Global Rise, Harvard Business Review, September 13th.

Faravelon, A., Frenot, S., & Grumbach, S. 2016. Chasing Data in the Intermediation Era: Economy and Security at Stake, IEEE Security and Privacy Magazine, Economics of Cybersecurity, Part 2.

Farronato, Chiara & Andrey Fradkin, 2018. "The Welfare Effects of Peer Entry in the Accommodation Market: The Case of Airbnb," NBER Working Papers 24361, National Bureau of Economic Research, Inc.

Foley, C.F., Hines, J.R., Mataloni, R.J., & Wesel, D. 2020. "Multinational Activity in the Modern World," the introductory chapter in the forthcoming Brooking volume edited by Bergsten et al. entitled Multinational Corporations in a Changing Global Economy.

Foerster, M. Data Center Investments in China: Cloud 9 for Foreign Investors?, JDSUPRA, June 16th.

Fox, M. 2020. 'COVID-19 Has Been Like Injecting Amazon with Growth Hormone': Here's What 4 Analysts Had to Say About Amazon's Earnings Report as $4,000 Price Targets Start to Roll In, Market Insider, July 31st. https://markets.businessinsider.com/news/stocks/amazon-earnings-wall-street-reacts-blockbuster-report-analysts-stock-price-2020-7-1029456482#

Furman, J, Coyle D, Fletcher A, McAuley, D, Marsden P, 2019. Unlocking Digital Competition. https://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel

Geradin, Damien, What Should EU Competition Policy do to Address the Concerns Raised by the Digital Platforms' Market Power? (September 30, 2018). Available at SSRN: https://ssrn.com/abstract=3257967 or http://dx.doi.org/10.2139/ssrn.3257967

Guan, W.Y. 2017. Internet Celebrity President Fight Hard for The Flourishing Age of Indonesia's Ecommerce, Business Weekly, November, 1565: 84-93.

Hall, B.H. 2005. Measuring the Returns to R&D: The Depreciation Problem, *Annales d'Economie et de Statistique* N^0 79/80, special issue in memory of Zvi Griliches, July/December. Also available as *National Bureau of Economic Research Working Paper* No. 13473, October, 2007.

Ho, L.R. 2020. Why Are Google's High Energy Consumption Data Centers Taiwan's New Sacred Mountain?, Business Weekly, September, 1712: 30-32.

Initium Media, 2020. The Era of Big Data. Publisher: Initium Media, Hong Kong.

Jones, Benjamin F. and Summers, Larry, 2020. A Calculation of the Social Returns to Innovation (September 2020). NBER Working Paper No. w27863

Ker, D. and E. Mazzini (2020), "Perspectives on the value of data and data flows", *OECD Digital Economy Papers*, No. 299, OECD Publishing, Paris, https://doi.org/10.1787/a2216bc1-en.

Kituyi, M. The digital divide is impeding development, https://unctad.org/en/pages/newsdetails.aspx?OriginalVersionID=1896

Kolodny, L. 2017. How Google Parent Alphabet Invests in Startups, CNBC, August 17. https://www.cnbc.com/2017/08/17/alphabet-google-start-up-investment-vehicles.html

Lev, B., and Radhakrishnan, S. 2005. The valuation of organisation 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. 2008. Global Sourcing in Innovation: Theory and Evidence from the Information Technology Hardware Industry." PH.D. Dissertation, University of California, Los Angeles.

Li, W.C.Y. 2015. Organization Capital, R&D Assets, and Offshore Outsourcing. 2015 ASSA/SGE

Sessions Proceeding Paper.
Li, W.C.Y., Nirei, M., & Yamana, K. 2019. Value of Data: There's No Such Thing as a Free Lunch in the Digital Economy, U.S. Bureau of Economic Analysis working paper, February.
Li, W.C.Y., & Chi, P.J. 2021. Online Platforms' Creative "Disruption" in Organizational Capital – The Accumulated Information of the Firm, Moon Economics Institute Discussion Paper, No. 1, August.
Lin, D. 2018. Amazon Gave Up its China's Cloud Servers to Compromise, the Epoch Times, October, 9th.
Lin, J. 2020. Amazon in India: Challenges Jio, World Journal, August 21st.
Lopez, M. 2016. How eBay Uses Big Data and Machine Learning to Drive Business Value, Forbes, October 4th, https://www.forbes.com/sites/maribellopez/2016/10/04/how-ebay-uses-big-data-and-machine-learning-to-drive-business-value/?sh=441cd6db1f35
Mattioli, D. Amazon Scooped UP Data From Its Own Sellers to Launch Competing Products, the Wall Street Journal, April 25th.
Mattioli, D., & Lombardo, C. Amazon Met With Startups About Investing, Then Launched Competing Products, the Wall Street Journal, July 24th.
Mohsin, M. 2020. 10 TikTok Statistics That You Need to Know in 2020, Oberlo, July 3rd. https://www.oberlo.com/blog/tiktok-statistics
Morton, F.S., Bouvier, P., Ezrachi, A., Juliien, B., Kimmelman, G., Melamed, A.D., & Morgenster, J. 2019.
Report on the Study of Digital Platforms, Chicago Stigler Center for the Study of the Economy and the State, May.
Munshi, N. 2020. Africa's Cloud Computing Boom Creates Data Centre Gold Rush, Financial Times, March 2nd.
Murphy, H. Facebook Unveils Global Digital Coin Called Libra, Financial Times, June 18th.
Nguyen D and M Paczos (2020), Measuring the Economic Value of Data and Data Flows.
Thomas Niebel, Fabienne Rasel & Steffen Viete (2019) BIG data – BIG gains? Understanding the link between big data analytics and innovation, Economics of Innovation and New Technology, 28:3, 296-316, DOI: 10.1080/10438599.2018.1493075

Poulsen, K., & McMillan, R. 2020. TikTok Track User Data Using Tactic Banned by Google, the Wall Street Journal, August 11th.
Reuters.com. 2000. Amazon to Hire 3000 People in South Africa for Customer Service Roles, Reuters Business News, June 18th. https://www.reuters.com/article/us-amazon-com-safrica/amazon-to-hire-3000-people-in-south-africa-for-customer-service-roles-idUSKBN23P1RE
Scott Morton, F, et al 2019. Stigler Committee on Digital Platforms. https://www.publicknowledge.org/wp-content/uploads/2019/09/Stigler-Committee-on-Digital-Platforms-Final-Report.pdf
SimilarWeb, 2020. Top Websites Ranking: Top Sites Ranking for All Categories in Vietnam, July,
https://www.similarweb.com/top-websites/vietnam/
Segulah. 2020. https://segulah.com/investering/hermes-medical-solutions-3/
Sender, H. 2020. How Competitive Is China's Tech Scene?, the Financial Times, August 4th.
Statista, 2020. Number of Facebook Users in China.
https://www.statista.com/statistics/558221/number-of-facebook-users-in-china/
Statista, 2020. Number of Facebook Users in the United States.
https://www.statista.com/statistics/408971/number-of-us-facebook-users/
Statistics Canada, Measuring investment in data, databases and data science: Conceptual framework. https://www150.statcan.gc.ca/n1/pub/13-605-x/2019001/article/00008-eng.htm
Stern, J. 2020. TikTokers Can't Quit TikTok. That's Why It's at The Center of A Global Business
Battle, the Wall Street Journal, August 3rd.
Sullivan, M. 2018. How Amazon's Startup Fund Is Betting on An Alexa-Everywhere Future, *Fast Company*, February 21st.
https://www.fastcompany.com/40476530/amazon-alexa-fund-startups-accelerator
The Economic Times, 2020. India Bans 59 Chinese Apps Including TikTok, WeChat, Helo, July 29th.
https://economictimes.indiatimes.com/tech/software/india-bans-59-chinese-apps-including-tiktok-helo-wechat/articleshow/76694814.cms
The Economist, 2020. ByteDance Is Going from Strength to Strength, April 18th.

https://www.economist.com/business/2020/04/18/bytedance-is-going-from-strength-to-strength
Thornhill, J. 2020. Smartwatches Call Time on The Swiss Industry, Financial Times, February 17th.
Tokopedia. Wikipedia. 2020.
https://en.wikipedia.org/wiki/Tokopedia#:~:text=In%202017%2C%20Tokopedia%20received%20%241.1,valuation%20to%20about%20%247B.
Tomiura, E., Ito, B., & Byeongwoo, K. 2020. Characteristics of Firms Transmitting Data Across Borders: Evidence from Japanese Firm-level Data, RIETI Discussion Paper Series 20-E-048, May.
Tzuo, T., & Weisert, G. 2018. Subscribed: Why the Subscription Model Will be Your Company's Future – And What to Do about It, Publisher: Portfolio.
UNCTAD 2021. Digital Economy Report (forthcoming).
Vernon R. 1966. International Investment and International Trade in the Product Cycle, *Quarterly Journal of Economics*, May: 190-207.
Wang, Y. 2020. Facebook and PayPal Will Invest in Gojek, the Indonesia Ridesharing Platform, the Wall Street Journal (Chinese Version), June 3rd.
Weber, S. 2017. Data, Development, and Growth, Business and Politics, 19(3): 397-423.
Wells, G., & Kubota, Y. 2019. TikTok's Videos Are Goofy. Its Strategy to Dominate Social Media Is Serious, the Wall Street Journal, June 29th.
Widowati, H. 2018. Amazon Heating Up E-Commerce Competition in Indonesia, September 18th,
https://en.katadata.co.id/inside/2018/09/18/amazon-heating-up-e-commerce-competition-in-indonesia
Williams, D. 2019. E-commerce in China: A market overview, October 7th,
https://www.retail-insight-network.com/features/e-commerce-in-china/#:~:text=The%20current%20state%20of%20e%2Dcommerce%20in%20China&text=And%20in%20June%20this%20year,data%20and%20research%20company%20eMarkter.
Yang, Y. 2020. Why TikTok Owner ByteDance Is No Huawei for Beijing, the Financial Times, August 3rd.
Yin, Y.H. 2018. The Most Successful Internet Company after Amazon – From Netherlands, Business Weekly, 1600, July, Taiwan.

Figure legends:

Figure 1: Global Data Flow vs. Combined Organizational Capital Stock of Dominant Online Platforms

Figure 2: The Information Pyramid

Figure 3: The Virtuous Circle of Data Sharing

Figure 4: Marriott's organization capital stock and the estimated data market size of the global hospitality industry

Figure 5: Type I: Amazon

---

[1] There are several potential revenue streams: advertising, selling the credit score of consumers, marketing analytics, providing risk management to financial service firms, and providing inventory management for third-party sellers.

[2] Foley et al. (2020) examine the data on the operations of multinational firms from the 1970s to 2017 and conclude that multinational firms tend to be based and operated in developed countries, though to a declining degree as middle-income Asian countries rise in importance.

[3] There are exceptions such as when U.S. firms produce in Vietnam, when a small part of process R&D may be produced by local engineers depending on the products. For IT hardware industry, though produced in China, the original engineers are from Taiwan (Li, 2008).

[4] However, Li (2020) did not find any effects on sales (or outputs), employment, and firm-level productivity due to the entry of online platforms during the period of 2002 to 2018 for the U.S. hospitality and transportation industries. The impacts of online platforms on firm-level productivity and employment may take longer to appear.

[5] The estimates presented here may include other drivers of organizational capital change but on the other hand are conservative because other firms like startups will also have demand for data.

[6] Since the 1970s, many economists have been attempted to estimate the depreciation rates of R&D assets. The Li and Hall (2020) model is currently the most widely adopted new methodology in the area of intangible research to estimate the industry-level and firm-level depreciation rates of R&D assets. Bronwyn Hall in private communication notes that the model can apply to other types of intangible capital as long as good-quality data on the investment of intangible capital and sales are available.

[7] The driving forces of the deprecation of intangible capital are competition and obsolescence (Li and Hall, 2020; Li and Chi, 2021; Hall, 2005).

[8] https://www.statista.com/statistics/247264/total-revenue-of-the-global-hotel-industry/

[9] Because Amazon does not disclose information on its data centers there is no public information available after 2015.

[10] At a 2018 OECD meeting, statistical agencies stated that Amazon's UK, Canada, and Japan offices all report that they do not make strategic business decisions for the firm, and that their main function mainly involves running warehouses to support Amazon's global logistics.

[11] Data source: Statistia and Williams (2019).

[12] India bans 59 Chinese apps recently, including top social media platforms, TikTok and WeChat (The Economic Times, 2020).

[13] Indonesia invested around 20% of national budgets in basic infrastructure in 2017 (Guan, 2017).