Data as a resource for the private sector

Main messages

1. Businesses are reaping tremendous value from both data created through businesses' economic activities and data shared by governments. Used as an input in data-driven decision-making, those data can spur innovation in products and services and reduce transaction costs, ultimately boosting productivity, export competitiveness, and growth.

2. Use of data in the production process of firms may help tilt the playing field toward poor people and underserved populations (who can trade across platforms and access free services) by reducing fragmentation in markets. However, it can also exacerbate domestic inequalities where foundational skills, infrastructure, and finance are not widely available in countries.

3. Use of data by businesses can also tilt the playing field away from poor countries, whose local enterprises may struggle to compete with large global players in part because of economies of scale and scope from data.

4. Although the use of data in the production process presents many opportunities to solve development challenges, policy makers should heed the risks this use presents for the concentration of economic power, patterns of inequality, and protection of the rights of individuals.

Creating value and solving development challenges through data-driven business models

For millennia, farming and food supply have depended on access to accurate information. When will the rains come? How large will the yields be? What crops will earn the most money at market? Where are the most likely buyers located? Today, that information is being collected and leveraged at an unprecedented rate through data-driven agricultural business models. In India, farmers can access a data-driven platform that uses satellite imagery, artificial intelligence (AI), and machine learning (ML) to detect crop health remotely and estimate yield ahead of the harvest. Farmers can then share such information with financial institutions to demonstrate their potential profitability, thereby increasing their chance of obtaining a loan. Other data-driven platforms provide real-time crop prices and match sellers with buyers.

For remote populations around the world, receiving specialized medical care has been nearly impossible without having to travel miles to urban areas. Today, telehealth clinics and their specialists can monitor and diagnose patients remotely using sensors that collect patient health data and AI that helps analyze such data.

Innovations like these herald the promise of business models that apply data to create new and better goods and services, helping to address development challenges in the process. Both private intent and public intent data are increasingly being used by firms to create value in the production process. At the same time, data are continually being produced as a by-product of economic activity, creating digital footprints that drive the data economy. With their growing capacity to collect, store, and process that data, businesses find that their ability to extract value from this data has been rising exponentially in recent years.

The COVID-19 crisis has created urgent demands for the private sector to adopt data-driven solutions to deal with the pandemic and increase resilience and productivity for recovery. Big Tech companies have been one of the few winners during the crisis as consumers purchase more goods and services online. As businesses shift toward recovery, the new reality will likely accelerate trends toward data-driven technologies that allow for automation and traceability in value chains.

For all their promise, however, the accelerating pace of these trends also comes with risks related to the concentration of economic power, greater inequality, and protection of the rights of individuals. The degree to which individuals can benefit from the data-driven economy—including consumers, entrepreneurs, and job seekers—will differ according to their access to finance, education levels, skills, and technology. In charting a way forward, policy makers—across all stages of development of their country's data-driven economy—should remain alert to these risks so that the use of data by firms contributes to broadly shared benefits.

The role of data in the production process of firms

The role of data in the production process can be conceptualized in different ways, depending on the specificities of the firms, industries, technologies, and types of data being considered. There is as yet no overarching theory or consensus on the role of data in the production process. The categories that follow summarize various ways of understanding the role of data in the creation of value by firms—as a factor of production, as a productivity enhancer, as a by-product, or as an output.

Data as a factor of production. For some firms, data are considered an input central to their business, essential to fulfillment of their core objectives. In this context, data have been referred to as a factor of production—on a par with labor, capital, and land—that is a primary determinant of output and productivity.[1] For example, many social media platforms are built around monetizing their users' data for advertising.

Data as a productivity enhancer. Data may also be conceptualized as a driver of total factor productivity (TFP). Increases in TFP reflect a more efficient use of factors of production often thought to be driven by technological change. Businesses use data along with various technologies to become more productive by improving their business processes, learning more about their clients and customers, developing new products, or making better data-driven decisions.[2] In this context, the addition of data to the production process makes the main factors of production more efficient, leading to better performance. According to one study, in the US health care sector the use of big data has been associated with a 0.7 percent increase in productivity growth per year.[3] Other studies have found that among 179 large publicly traded US firms the adoption of data-driven decision-making has led to an increase in productivity of 5–6 percent.[4]

Data as a by-product of the production process. Data are often passively created as a by-product of economic activities. For example, call detail records (CDRs) are a by-product of telephone usage. Observed data on consumers' browsing and buying patterns are a by-product of online e-commerce. Data created in this way can be used in the production of new products or services, either by the firm that produced the original data or by other firms with which the data are shared, such as under commercial arrangements. For example, e-commerce platforms use data created as a by-product of transactions on their platform to improve their product offerings; credit card companies sometimes sell their transaction data for a specific location to firms involved in tourism in that location; and new firms use CDRs for commercial purposes, including analytics and advertising.[5]

Data as an output. For some firms, data are the primary output of the production process. Examples are data intermediaries, including rating services such as Nielsen; pollsters such as Gallup; and data aggregators such as dataPublica.[6] These data are then used either by other firms in their production processes or by government in policy making.

In all cases, data have a role in creating value for the economy, but the way in which data play into the production process differs by context.

Pathways to development

Whether the use of data in the production process is conceptualized as a factor of production or a driver of productivity, its transformative effects on development can be summarized by four channels:

1. *Quality improvements in existing products and services.* This channel includes the use of data-driven decision-making to provide consumers with better health diagnostics, better credit scoring, better search results, and more personalized product recommendations.
2. *Cost reduction in delivering products and services.* Data and analytics can reduce the costs of delivery, which can then lower prices (subject to markets being sufficiently competitive). For example, better credit scoring can reduce the cost of delivering loans and lead to lower interest rates on loans. Sensor-based agricultural devices and platforms that take and analyze soil readings can inform farmers how much fertilizer they should apply, which should reduce wastage and costs.
3. *Greater innovation in development of new products and services.* Examples include the development of new financial products, smart contracts and supply chain tracking services, new products that rely on applications such as online maps or translation, and new consumer goods based on analysis of purchasing trends.
4. *More effective intermediation and lower transaction costs.* Platform firms can help solve market failures and lower the entry and transaction costs for firms that connect to those platforms. This happens in part by reducing information asymmetries, thereby increasing trust in those firms. Distributed ledger technologies (DLTs) not only can reduce transaction costs but also enhance trust through secure transactions. Better intermediation can disrupt traditional market structure and reduce the market power of intermediaries, particularly in sectors such as agriculture where they have traditionally played a central role in the value chain.

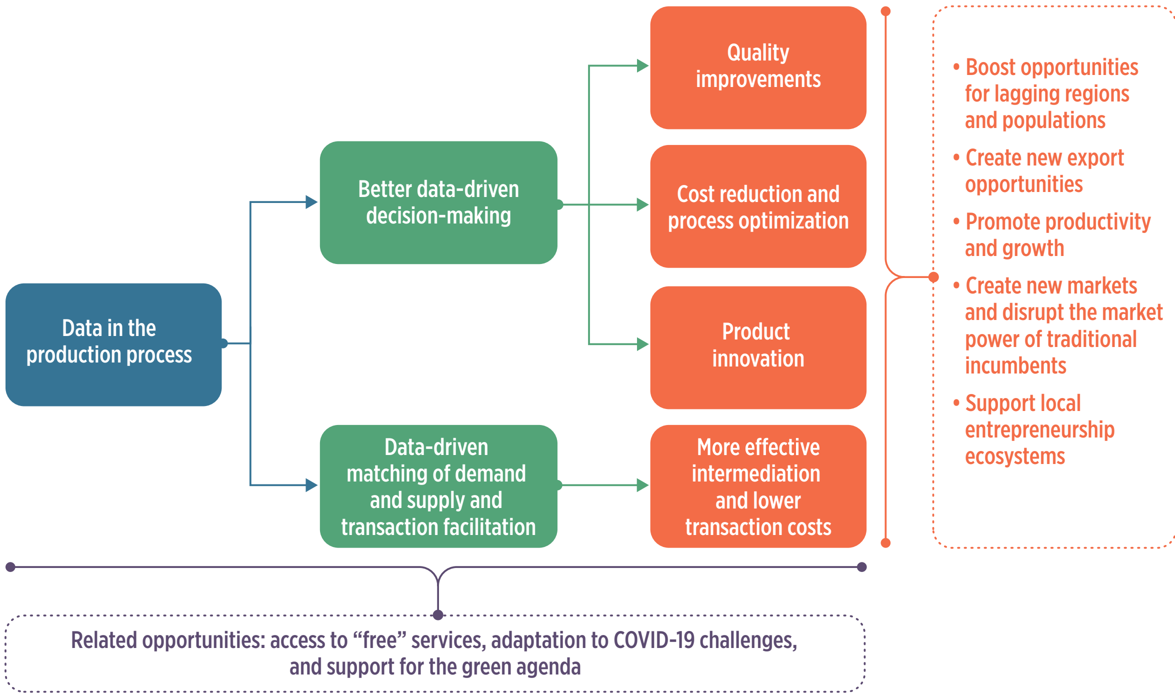
These four channels to increasing the impact of data on development are driven by two key effects. First, analytics applied to data can reveal patterns that *allow better data-driven decision-making*. Second, data can help to *facilitate transactions, including by matching the suppliers of goods and services with those who demand them*. In this way, the use of data can help overcome market failures, with positive effects on productivity, growth, jobs, and welfare (figure 3.1).

Data-driven businesses and the technologies that help them create value

Data-intensive analytics can be used to discover new insights, enhance decision-making, and optimize processes. When data are characterized by the "3 V's"—volume, velocity, and variety—they can serve as inputs to big data analytics. Such analytics typically require new methodologies and technologies to enable enhanced decision-making (box 3.1). This chapter focuses on the development impact of business models that use data-intensive technology or analytics as their key value drivers, whether they are technology firms (the providers of data-intensive technological solutions) or traditional firms and entrepreneurs (the adopters of data-intensive technologies).

Firms may use various data-driven technologies by themselves or in combination. A key business model that has emerged using data-intensive technologies are *data-driven platform businesses*, which use data, along with AI/ML and other analytics, to intermediate between distinct user groups to match supply with demand. By overcoming informational asymmetries and reducing search costs, these businesses facilitate market exchanges and generate more data on users and their behavior. Some may also use a combination of other technologies. For example, the platform GrainChain uses DLT to broker secure transactions

Figure 3.1 The role of data in the production process: Pathways to development



Source: WDR 2021 team.

Box 3.1 Technologies and methods that support data-driven decision-making and intermediation

Technology that supports data-intensive analytics: artificial intelligence, including machine learning
Artificial intelligence (AI) and machine learning (ML) can help firms analyze their data with less manual effort. AI is the development and use of any device that perceives its environment and takes actions that maximize its chance of success of reaching a defined goal (including learning and adapting to its environment). It is not a single technology but a family of technologies. Machine learning is one application of AI. The algorithms that underlay AI rely on inputs of large amounts of data to learn and produce accurate and valuable insights. Based on adoption patterns, studies predict that firms responsible for about 70 percent of economic output will have adopted at least one type of AI technology by 2030.[a]

Data-intensive analytic applications and big data analytics sometimes require that data be processed in different formats and distributed across different locations. These may include cloud computing, bio-inspired computing, or quantum computing. They also require the capacity to store big datasets and to clean them to correct inaccuracies.

Technology that collects data and actions insights from analytics: smart devices and devices connected through the Internet of Things (IoT)
Devices include sensors and monitors that generate data. Smart devices rely on these "machine-generated" data to improve their operations, often using AI. Devices are increasingly being connected to the IoT, which allows them to receive and send data from and to other IoT devices on ground moisture, climate and air quality, individuals' health metrics, firm asset performance, and the movement of goods through supply chains. IoT and machine-generated data from devices are poised to multiply exponentially the data generated by businesses, with potential for development in agriculture, health, manufacturing, and transportation (such as driverless vehicles). IoT devices already exceed the number of internet users and are forecast to reach 25 billion by

(Box continues next page)

Box 3.1 Technologies and methods that support data-driven decision-making and intermediation *(continued)*

2025, with the introduction of fifth-generation (5G) wireless technology.

Technology that creates transparency and trust in data records: distributed ledger technology, including blockchain

Distributed ledger technology (DLT) is a distributed database in which data are recorded, shared, and synchronized across the nodes (or devices) of a network. Blockchain is a type of DLT whereby information is consolidated into "blocks" that are linked in a way in which they can add information layers to the ledger, which cannot be changed (in an "append-only" fashion). Blockchain records transactions, tracks assets, or transfers value between two parties in a verifiable and permanent way without the need for a central coordinating entity. Because everyone participating in the blockchain can see all transactions, the technology engenders peer-to-peer trust and has several applications, including enabling payments, smart contracts, supply chain tracking, and resolving data protection and security issues in the IoT.

a. MGI (2018).

between buyers and sellers of agricultural commodities, while employing Internet of Things (IoT) devices to accurately measure variables such as commodity weight. In those platform businesses that earn a significant proportion of their revenue from advertising, data collected through the platform are used to inform that advertising. Platform models are a key focus of this chapter and of many of the economic policy issues raised in chapter 7 because of their importance to low- and middle-income countries.

The extent to which data-intensive technologies can be deployed relies on the presence of key infrastructure, most fundamentally network coverage. There is also the challenge of bringing more people online, especially in countries with a gap between the number of people who have access to networks and those who are online. This gap is a function of affordability, the existence of local content, and digital skills (see chapter 5). As more people and devices come online and data usage matures, the network capacity needed will grow, requiring sufficient spectrum to be made available for mobile use—especially in lower-income countries where mobile is the predominant technology. Although 4G technology is sufficient for many IoT uses, 5G will be needed for those uses that require ultra-reliability and low latency such as smart energy grids and autonomous vehicles. Reliability of connection is important for DLT applications that must keep a reliable and consistent record of data. Storage and analysis of the data generated through IoT devices and platform business models depend on cloud computing (remote storage and processing infrastructure) and the ability to transmit data over the internet to data centers either locally or abroad. Beyond data infrastructure, most technology applications require a suite of other foundational systems to create value, including reliable payment systems and logistics networks, transport infrastructure, and address systems. Data infrastructure policy is discussed further in chapter 5.

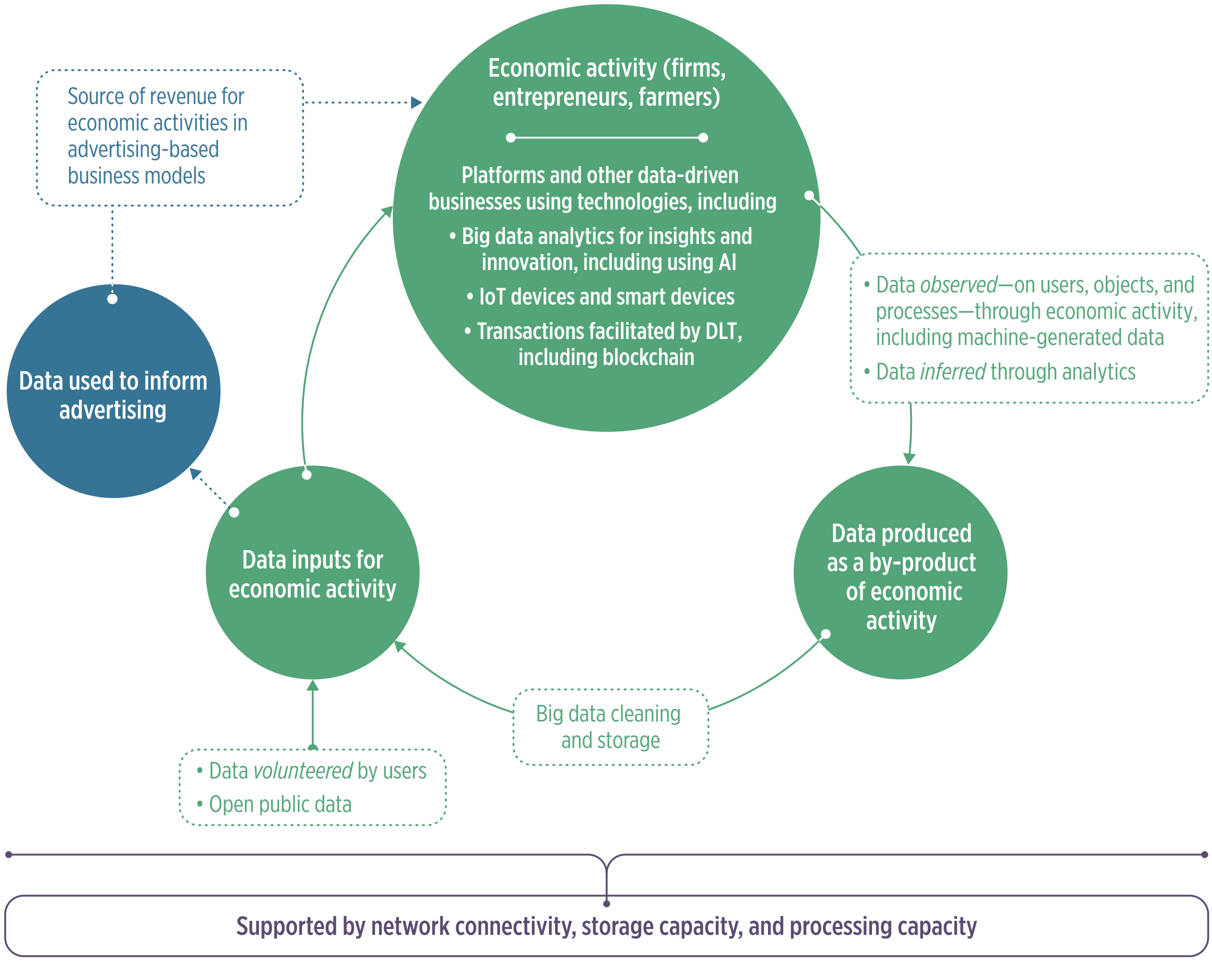
Figure 3.2 summarizes how data are used as an input to and produced as a by-product of economic activity. It illustrates how data created through economic activity can be used as an input to either the same economic activity or new activities.

Focus on platform firms in low- and middle-income countries

Platform businesses, one of the most ubiquitous and transformative data-driven models today, reduce transaction costs and alleviate market failures.[7] Ranging from start-ups to businesses operating at scale, they are a mix of both locally grown and foreign firms, and they are expanding across low- and middle-income countries. More than 300 digital platforms headquartered in Africa were active across major Sub-Saharan African economies as of 2020.[8] In Asia, a study looking at local platforms that had reached scale identified 62 major local platforms with an individual market capitalization of at least US$800 million as of 2016, half of which were located in China.[9]

The diversity of new platforms is evident in recent research examining both start-ups and scaled platforms. At least 959 platform firms have established a physical presence in a sample of 17 low- and middle-income countries[10] from all regions across four sectors that are important for jobs or economic productivity: e-commerce, transport and logistics (including both freight and passenger transport), agriculture, and

Figure 3.2 The role of data in economic activity



Source: WDR 2021 team.

Note: AI = artificial intelligence; DLT = distributed ledger technology; IoT = Internet of Things.

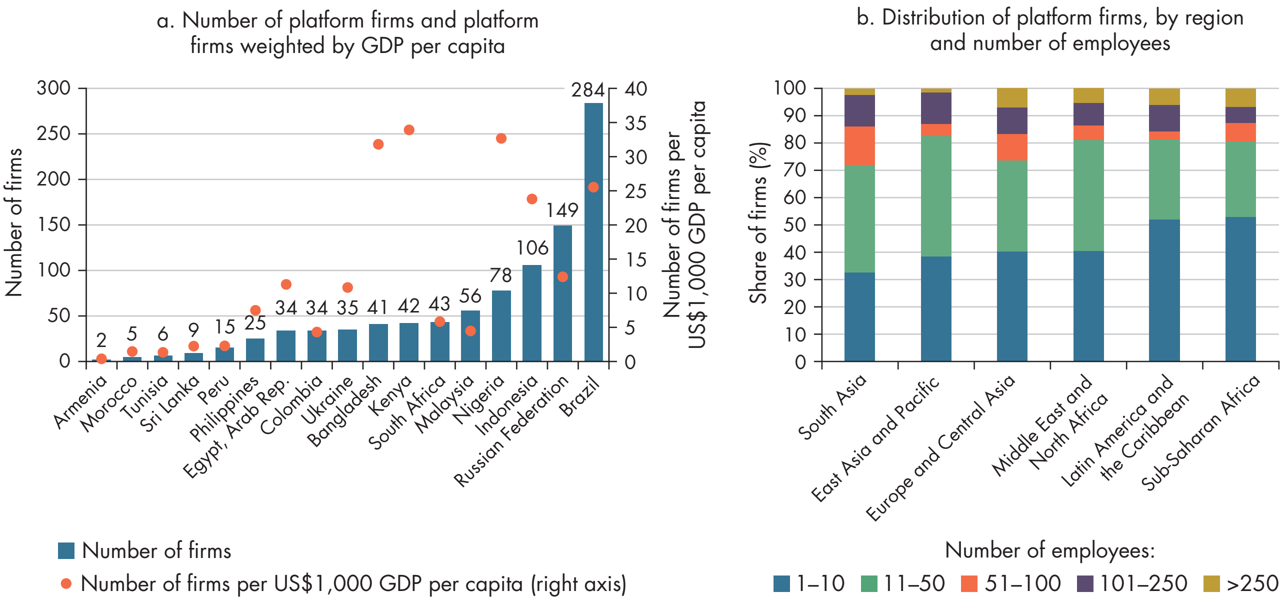
tourism.[11] In the sample, Bangladesh, Brazil, Indonesia, Kenya, and Nigeria have relatively high numbers of platform firms when controlling for gross domestic product (GDP) per capita (figure 3.3, panel a).

Across the countries in the sample, most platform firms are recent entrants—55 percent were established in the past five years.[12] Only 11 percent of firms were established more than 10 years ago. Firms also tend to be small—over 80 percent have 50 or fewer employees, and almost half (47 percent) have 10 or fewer (figure 3.3, panel b). Most firms have remained active (defined as having an active and up-to-date online presence) since they were established; the average share of firms currently active across regions is more than 80 percent. Sub-Saharan Africa is an outlier: nearly half of its firms appear to be inactive.[13]

E-commerce has the highest share of platform firms in 82 percent of countries in the sample, with the highest shares in South Asia and the Middle East and North Africa and the lowest in Europe and Central Asia. The agriculture sector tends to have the smallest share of firms across regions, with the exception of Sub-Saharan Africa. The importance of e-commerce in the data economy is also reflected in web traffic.

Although local data-driven firms are on the rise in low- and middle-income countries, foreign-headquartered firms have a significant presence, underscoring the global nature of the data-driven economy. Their presence is also a reminder that the platform economy is still nascent in lower-income countries relative to high-income economies (partly due to issues around trust, lack of digital skills, and lack of access to finance). Of the top 25 websites in terms of traffic in the 17 low- and middle-income countries sampled, 59 percent belong to firms with foreign headquarters on average[14]—however, the figure varies across countries (figure 3.4). Although the presence of firms from high-income countries in lower-income countries is

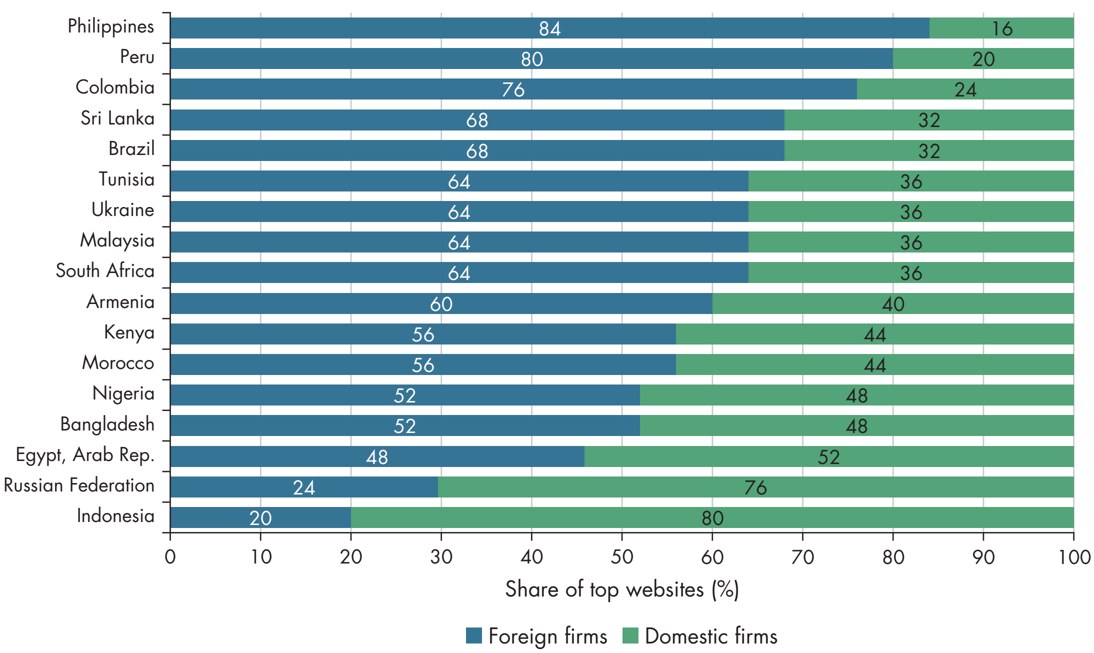
Figure 3.3 Platform firms are numerous in some lower-income countries but tend to be small



Sources: Nyman and Stinshoff (forthcoming), based on information from Crunchbase, Crunchbase (database), https://www.crunchbase.com/; World Bank, Digital Development (database), https://www.worldbank.org/en/topic/digitaldevelopment; Dow Jones and Company, Factiva (database), https://professional.dowjones.com/factiva/; Thomson Reuters Foundation, "Inclusive Economies," http://www.trust.org/inclusive-economies/; Alexa Internet, "The Top 500 Sites on the Web, by Country" (accessed April 2020), https://www.alexa.com/topsites/countries; World Bank, World Development Indicators (database), https://datatopics.worldbank.org/world-development-indicators/.

Note: Panel a shows the number of platform firms and platform firms weighted by gross domestic product (GDP) per capita in selected low- and middle-income countries The total sample of platform firms is 959. Per capita GDP is in constant 2010 US dollars for 2019. Panel b shows the share of firm sizes in terms of number of employees by region in a sample of 595 active platform firms.

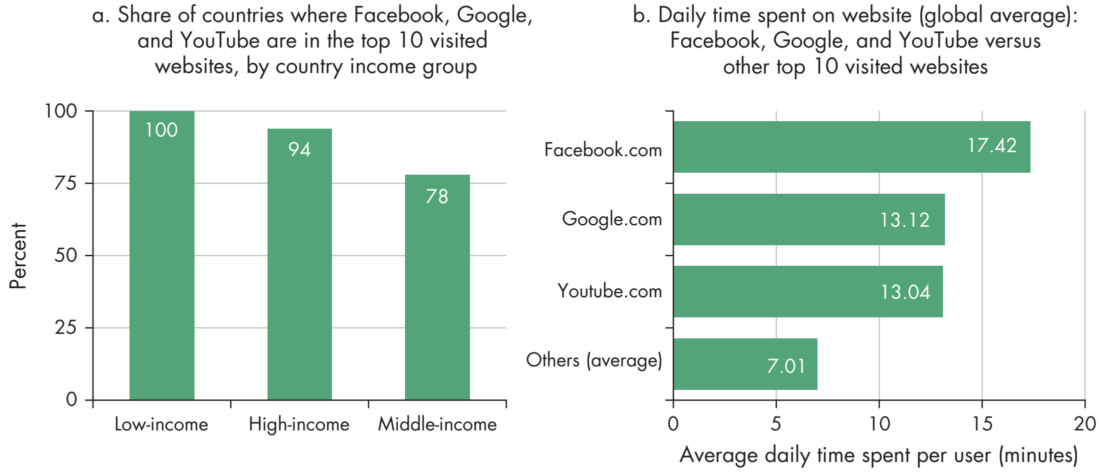
Figure 3.4 The importance of domestic versus foreign-headquartered firms differs across countries as indicated by firm share of top websites



Source: WDR 2021 team, based on Alexa Internet, "The Top 500 Sites on the Web, by Country" (accessed April 2020), https://www.alexa.com/topsites/countries.

Note: The figure shows the percentage of websites for firms with foreign headquarters versus domestic headquarters among the top 25 websites per country based on traffic. Headquarters is understood to be the global headquarters, not the domestic or regional office. Total sample size is 425 websites.

Figure 3.5 Users visit and spend more time on Facebook, Google, and YouTube than other websites



Source: WDR 2021 team, based on data from Alexa Internet, "The Top 500 Sites on the Web, by Country" (accessed April 2020), https://www.alexa.com/topsites/countries.

Note: Sample of 1,270 websites (top 10 websites in 127 countries).

widespread, the opposite is not true. Only 15 percent of digital firms headquartered in Sub-Saharan Africa operate outside the region, and the majority of those have expanded to the Middle East and North Africa.[15]

The leading global platforms are highly relevant to the digital ecosystems of lower-income countries and their citizens, particularly for online search and social media. Google, YouTube (which is owned by Google), and Facebook are among the top 10 most visited websites in 62 of 77 low- and middle-income countries (figure 3.5, panel a). These platforms also have the highest average daily time spent on the site per user globally (figure 3.5, panel b). In online markets where firms compete for the attention of viewers, such popularity can significantly intensify these platforms' market power in advertising (which is, in turn, important for suppliers of other products) and increase the amount of data being collected about users. Google's Next Billion Users initiative is specifically aimed at developing products and services for lower-income countries. Facebook has launched an app aimed at providing free data in lower-income countries. WhatsApp (owned by Facebook) is by far the most used mobile application globally in terms of time.[16] Because of the global nature of these firms, dynamics in overseas markets that affect the strategies and policies of these large platforms will have repercussions for those in low- and middle-income countries.

Data traffic over the internet is also highly concentrated in a few companies. Six US companies generate more than 40 percent of the world's internet data flows (figure 3.6, panel a). Across the top 25 websites (by traffic) in the 17 sampled countries featured in figure 3.4, some 60 percent is owned by five firms headquartered in the United States (Google, Microsoft, Facebook, Verizon, and Amazon). Significant non-US parent companies include Naspers (headquartered in South Africa), Alibaba (China), and Jumia (which has its operations largely in Nigeria)—see figure 3.6, panel b.

Data inputs for economic activity

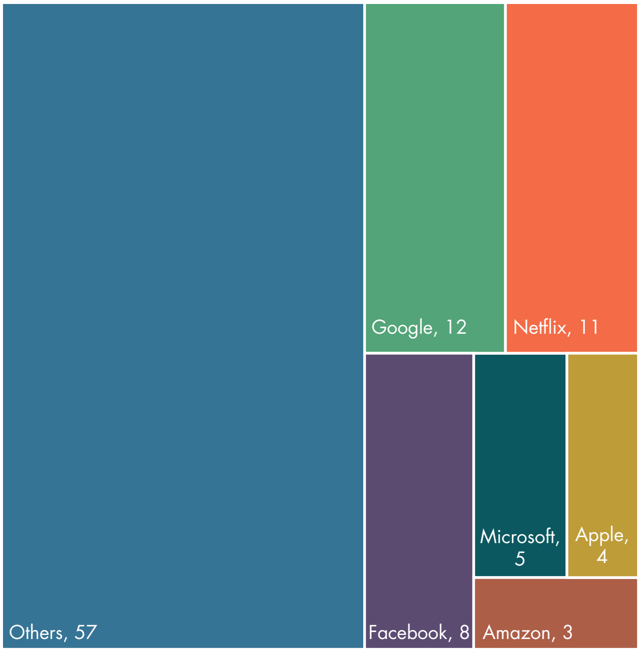
The "digital footprint" and data collection by firms

Everything a digital user does leaves a trail, whether it is making a phone call, sending a text, conducting an online search, posting on social media, or making a digital transaction. The digital footprint of an individual or business is their collection of traceable digital activities and communications on the internet or other digital media. Data collected through devices—particularly IoT devices—can also capture insights on individuals and firms. For firms, such insights are gained from data on throughput and efficiency, spare capacity, and asset quality, among other things. For individuals, they typically involve health and biometric data.

Digital footprints can be actively created when a user makes a choice to share information, such as by posting on social media or volunteering information to register for services. Or they can be passively

Figure 3.6 Internet traffic in low- and middle-income countries is concentrated in several US-based firms

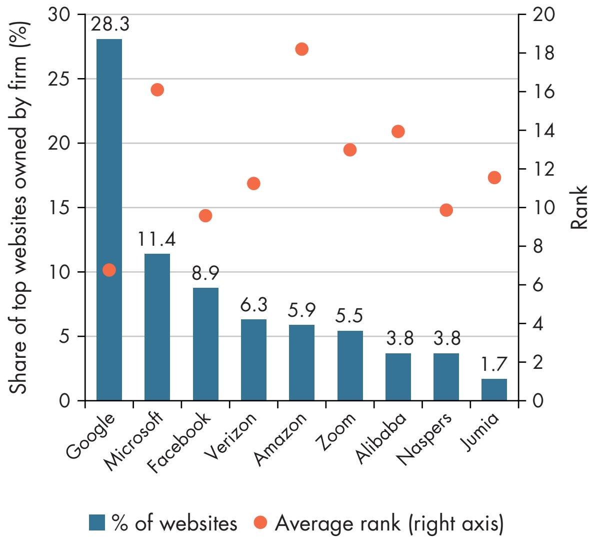
a. Share of global internet traffic by firm, 2018 (% of total)



Source: Sandvine 2019. Data at http://bit.do/WDR2021-Fig-3_6_a.

Note: "Others" refers to file sharing, marketplace, security, and virtual private network (VPN) messaging, as well as cloud and audio streaming.

b. Ten most visited parent firms among the top 25 websites



Source: Nyman and Stinshoff (forthcoming), based on information in Alexa Internet, "The Top 500 Sites on the Web, by Country" (accessed April 2020), https://www.alexa.com/topsites/countries.

Note: The panel shows the percentage of the top 25 websites owned by the listed parent organizations in the sample. Traffic rank is from a low of 1 (more traffic) to a high of 20 (least traffic). Total sample includes 425 websites from 17 low- and middle-income countries. Ownership is defined as majority shareholding.

created, when data are left behind as a by-product of other activities such as an Internet Protocol (IP) address, search history, or internet clicks. Firms typically collect both active and passive data. Often, this information is collected at exceedingly high frequency and microgranularity.

Big data and ML algorithms enable firms to draw inferences about the characteristics of individuals (such as attitudes and socioeconomic status) and other businesses (such as performance, capacity, and reputation). In low- and middle-income countries, digital footprints are best known for their ability to predict repayment behavior. However, applications also extend to the development of new products and improvements in service delivery across the economy, to the development of analytical tools for businesses, and, importantly, to the creation of consumer profiles that can be used to sell advertising services—a practice that ultimately subsidizes many of the "free" products that consumers use today. A number of risks have been identified and concerns raised about these methods and applications. These are discussed later in the chapter.

Data collected through mobile phone use has been one of the fastest-growing sources of user information and behavior. CDRs held by mobile network operators (MNOs) contain certain data on every call and text made, including the telephone number of the caller and receiver, the date and time of the interaction, and the associated cellphone tower. This information is primarily collected for billing purposes but can be used as well to identify the behavior, mobility patterns, and social networks of users. MNOs can also track data on use of value-added services, internet services, and mobile money transactions.[17]

Data-driven firms—including e-commerce, online search, and social media firms—produce, in addition, data on behavioral patterns that can be detected by noting the time, frequency, and extent of transactions or communications made. Several variables that can help predict economic status are available to platform firms, including the device type of their users (desktop, tablet, mobile), the operating system (Windows, iOS, Android), and the channel through which a user arrived at the firm's home page. For example, having an iOS device consistently correlates with being in the top quartile of the income distribution in the United States.[18] A study in Germany found that the time of day that purchases were made on an e-commerce site was predictive of a consumer's self-control and repayment behavior. People who made purchases between noon and 6 p.m. were half as likely to default on their payment for the products bought as those who made purchases between midnight and 6 a.m. Consumers coming from a price comparison website were roughly half as likely to default as customers directed to the site via search engine ads, consistent with research on impulse shopping.[19]

Finally, firms collect data that are generated as a result of phenomena that are not attributable to a specific individual or business. These data are often generated by machines or devices and can include data on traffic, weather and climatic conditions, and network usage in the case of computing or network industries. This Report highlights many examples of using this type of data for development purposes.

The use of open public intent data by businesses

Public intent data are also used by businesses for commercial purposes, particularly where those data are nonpersonal and where there are positive spillovers from the private use of such data. The use of open public intent data by private firms is prevalent in advanced economies with advanced national data systems, although there are also examples from low-income and emerging economies. Spotlight 3.1 discusses these uses in greater detail.

The positive development impacts of data used in the production process

This section presents examples of potential positive development impacts that data used in firms' production processes can facilitate (following the third development pathway of data generated by private firms described in chapter 1). It then turns to the risks stemming from the use of data by firms that policy makers need to consider. These examples are included to help set out what the current production possibility frontier might look like going forward. Although many of the examples may be outliers, they can indicate what may be possible and what countries may want to aspire to.

Boosting opportunities for lagging regions and populations by reducing market fragmentation. Data-driven business models can lower entry costs to markets and provide new opportunities for small firms and low-income households.[20] Firms in less populated areas can expand their access to markets through data-driven platforms, which match sellers and buyers, and through the logistics services of e-commerce platforms, which provide support in bringing products to market from more remote areas. Costs associated with distance are as much as 65 percent lower for online vendors active on global e-commerce platforms, compared with those for offline vendors.[21] Lower-income countries could reap substantial benefits from such e-commerce platforms: the cost-reduction effects of platforms tend to be greater for exporting countries that are unknown or less trustworthy to consumers (as measured by corruption indexes).[22]

In China's rural (and traditionally lower-income) Taobao Villages, where annual e-commerce transaction volumes exceed 10 million yuan and at least 10 percent of households engage in e-commerce, rural households trading goods on the Taobao platform have significantly higher incomes and higher income growth than those that do not.[23] These data are suggestive of the large benefits of data-driven business models in lagging regions, but the evidence base is just beginning to develop, and not all signs are encouraging. For example, a recent randomized trial in China that connected rural villages to e-commerce found little evidence of income gains for rural producers and workers. Understanding what factors led to Taobao success and meager gains for other Chinese villages is a crucial part of the future research agenda.

Creating new export opportunities. Not only can entrepreneurs market their goods remotely, but intangible data-enabled services now flow across borders. The boom in data-enabled services creates opportunities for new entrants in global trade and may foster economic growth for countries traditionally lagging in access to global markets. For example, the rise in Indian services exports has been associated with gains in per capita income and a decline in urban and rural poverty head count ratios.[24] The Indian digital services industry has also created employment, especially for women and in smaller cities with populations of about 1 million or less, which can help bridge economic and social inequalities.[25]

The Philippines is another important beneficiary, exporting an estimated US$23 billion in services enabled by information technology (IT)—equivalent to almost half of the country's merchandise trade exports and more than double its total agricultural exports.[26] In Africa, Senegal, a low-income country, boasts a dynamic digital services sector focused on business services as well as apps and software solutions geared toward regional markets. Mauritius has developed a fast-growing export-oriented digital business services industry, diversifying its services-centered economy away from the country's tourism sector.[27] Such expansion into digital services has proven important in light of the tourism collapse provoked by the COVID-19 pandemic. Studies show that trade in services, in particular IT-enabled business services, is not as prone to sudden collapses as other forms of global trade.[28]

Promoting productivity and growth. Although evidence from lower-income countries is scant, across four industries (hotels, restaurants, taxis, and retail trade) in 10 member countries of the Organisation for Economic Co-operation and Development (OECD)[29] the average service provider enjoyed bigger increases in the overall productivity of labor and capital in countries with relatively high online platform development between 2011 and 2017.[30] Increased e-sales activities accounted for 18 percent of the labor productivity growth in 14 European countries from 2003 to 2010.[31] There is also some evidence from Europe of a smaller productivity gap between large and small firms in the sectors that use online sales most.[32]

Machine learning could fundamentally revolutionize innovation, bringing data to the core of the growth process.[33] Empirical evidence on 18,000 US manufacturing plants between 2005 and 2010 finds that firms with more intensive data usage are significantly more productive due to not only technology adoption but also data-driven decision-making.[34] Globally, AI could deliver additional economic output of about US$13 trillion between 2018 and 2030 (netting out competition effects and transition costs), boosting global GDP by about 1.2 percent a year, according to a 2018 modeling exercise.[35]

Of course, not all countries will be affected similarly, and the evidence base for low- and middle-income countries needs to be developed. The focus in this chapter on data-driven business models and productivity enhancement through data reveals an imbalance toward case studies from higher-income countries. That imbalance is itself a sign of the lopsided distribution of benefits deriving from data.

Creating new markets and disrupting the market position of traditional incumbents. E-money platforms, among others, have challenged traditional banks and other service providers in transferring funds to and from accounts. For example, M-Pesa reached 9.5 million customers within its first three years in operation in Kenya, a country with only 8.4 million bank accounts.[36] Electronic freight exchanges such as uShip and Mober that match carriers with cargo holders have triggered global logistics providers such as Schenker and DHL to develop their own digital exchanges. Incumbent taxi providers who were protected by fixed license caps are now exposed to competition from ride-hailing apps, which can improve inclusion of some parts of the population. For example, in Mexico City the proportion of female drivers with Uber (5 percent) is higher than in Mexico's taxi industry (0.5–2.5 percent).[37] A 2017 study of 2,000 firms in 60 countries found that digital entrants boost the size of an industry by both realizing latent demand and taking market share from incumbents.[38] Indeed, the study showed that the profits of incumbents fall significantly in response to competition from digital entrants, and the slowest-growing incumbents are the most affected. These effects should translate into welfare benefits for users through more affordable products and greater innovation, although the benefits are less likely felt by the lowest income groups because they do not participate in these markets as consumers or suppliers.

Supporting local entrepreneurship ecosystems. An ecosystem is an interconnected set of services accessed through a single integrated experience. Facebook, for example, enables users to shop, make hotel bookings, message contacts, read the news, and chat with a doctor—all with different firms but through a single interface. Successful data-driven firms often spark new business models through complementary products and aftermarkets. Such local ecosystems, consisting of symbiotic and interdependent firms, frequently rise up around leading multinational platforms. The leading global platforms are therefore highly relevant for the digital ecosystems of countries and their citizens.

Ecosystems built around larger firms can enable lower-income economies to build digital capabilities, especially because they integrate data across a series of services to increase the scale and scope of datasets. For example, Apple's Healthkit platform offers Apple device users the option to share their health and activity data across applications on their smartphones. This integration allows researchers, hospitals, and developers of health care and fitness apps to access valuable data to inform patient care, marketing, and product development. The development of a healthy ecosystem relies on provision of access to data and systems.

Related opportunities arising from data-driven business models

Three related opportunities can arise from data-driven business models that are not linked directly to the four channels discussed earlier. Evidence is just emerging, but it indicates that these models could be important for lower-income countries.

Providing nominally "free" services to consumers. Nominally "free" or "zero price" services have become an integral part of our lives. Free messaging services, video communication, social connection tools, search engines, map services, storage, and translation and payment apps are now commonplace and increasingly being developed with low-income consumers in mind. Google is rolling out scaled-down search and e-mail apps for low-end smartphones, as well as voice search in various dialects to overcome literacy challenges. Free services also offer important inputs for other data-driven businesses. Digital start-ups often rely on integration with digital payment systems, cloud storage, and online analytical tools. For transformative business models that match and connect users in different geographic locations, online map services are a crucial input.

The welfare gains of nominally "free" digital goods are substantial. However, they are likely to be underestimated because they are not captured in GDP—they have a nominal price of zero. A recent study found that including the welfare gains from Facebook would add between 0.05 and 0.11 percentage points to GDP growth per year in the United States.[39]

"Free" services reflect the very low marginal cost of replicating and distributing data and certain digital services. But they are ultimately made possible on a large scale because firms can monetize data through advertising and data sales, thereby giving rise to the idea that users in fact "pay with their data." This approach is reflected in the revenue structure of some of the tech giants. Mobile advertising made up 84 percent of Google's total revenue in 2019,[40] while Alibaba earns more than half its global revenues from advertising.[41]

Both Google and Facebook have offered free internet access in lower-income countries. Facebook's schemes have been criticized for mining the data of low-income users while initially defying net neutrality rules and offering access to only a limited set of sites.[42] Meanwhile, Google attempted to monetize its free Wi-Fi service (rolled out in nine middle-income countries) by showing ads to users, but it recently discontinued these services because they have proven unprofitable.[43]

More limited opportunities in low- and middle-income countries for monetizing data may therefore limit the ability of firms operating locally to offer free services. Facebook's average revenue per user in the United States and Canada was US$41.41 in 2019, whereas it was US$2.48 in all countries except the United States, Canada, and European and Asia-Pacific countries. Firms focused on lower-income countries may find it difficult to replicate the free services offered by firms that operate globally and can cross-subsidize their operations with global advertising.

Adapting to new ways of doing business because of the COVID-19 pandemic. Urgent demands have surfaced for the private sector to adopt data-driven solutions to deal with the pandemic and increase resilience and productivity for recovery. Firms will also need to increasingly invest in the transparency and traceability of value chains, increase their reliance on automation in the production process, and make more precise predictions about their demand and input supply that anticipate disruptions. Data-driven technologies will play a critical role in helping firms adapt. Smart connected devices and robots that automate previously manual processes while collecting and analyzing data will serve as a key input in this reengineering of business processes during the recovery. AI that can predict consumption and production trends, combined with platforms that provide matching through data analysis, may also create on-demand labor forces.

Although the intensifying adoption of data-driven business models can be an opportunity, these trends also hold risks for the international competitiveness of those countries not at the technological frontier, with implications for jobs and inequality.

Impacts on the green and sustainability agenda. The increased use of data-intensive technologies contributes to global carbon emissions. And yet these technologies can also help firms better manage their environmental footprint and become sustainable, while allowing sectors such as agriculture to adapt to climate change. By improving the efficiency and traceability of supply chains and production processes, these technologies can reduce waste, enable circular solutions, promote sustainable sourcing of inputs, and empower consumers to make more environmentally responsible decisions. By making energy systems more efficient (including through automated tracking of energy use), they can facilitate the adoption of renewable energy through better management of performance. Data-driven farming can help farmers adapt to climate change while rationalizing use of harmful inputs. However, the net impact of such technologies on the environment will depend on several factors, including responsible actions by consumers and the decarbonization of the energy sector.

How use of data in the production process is transforming sectors

New business models that use data to drive value are springing up in low- and middle-income economies. The data and technology that can be most transformative depend on the types of market failures that need to be solved and the development channels that are possible. This differs across sectors.

Finance

Some 1.7 billion adults worldwide did not have a bank account as of 2017.[44] At least 200 million small firms in low- and middle-income countries have unmet credit needs estimated at US$2.2 trillion.[45] Several market failures are to blame. First, the high cost of traditional banking relative to the low-value transactions and balances of low-income individuals makes it less viable or attractive for traditional banking to serve this market segment. Second, information asymmetries between financial institutions and low-income borrowers make it difficult to assess credit risk, thereby limiting the supply and raising the price of credit. Finally, formal financial services lack relevant products and services for low-income users. Digitization and data analytics can help overcome these challenges to make services more accessible, affordable, and secure.

Alternative credit scoring algorithms. Financial service providers are increasingly adopting alternative credit scoring techniques that take advantage of users' digital footprints to train ML algorithms to identify, score, and underwrite credit for individuals who otherwise lack documentation of their creditworthiness.

Two early movers that have achieved scale—Lenddo (Philippines) and Cignifi (operating in Africa, Asia, and Latin America)—use data that consumers volunteer about their cellphone use patterns, digital transactions, and social media and web browsing activity to build algorithms that map behavioral patterns and score the creditworthiness of borrowers.

Payment and transaction histories have also enabled e-commerce firms to move into lending. Ant Financial's MYbank app links directly to users' Alibaba transaction data to score and extend credit.[46] Amazon's small business loan operation (which operates in China, India, and other countries) relies on a seller's sales performance on Amazon to decide whether to extend credit. Destácame, a Chile-based alternative credit scoring start-up and the first of its kind in Latin America, uses data on utilities payments for its credit scoring.

Psychometric tests, which assess the abilities, attitudes, and personality traits of individuals, are also being used to screen borrowers. LenddoEFL provides financial institutions with psychometric tools that analyze applicants' answers on an online quiz, including factors such as how long it takes applicants to answer and how they interact with the web interface. LenddoEFL claims to have made more than 12 million credit assessments through more than 50 client financial institutions around the world.[47]

Notwithstanding the opportunities these approaches offer, observers have raised concerns that using algorithms in this way can discriminate against individuals and reinforce existing racial, gender, and economic inequalities. Lenddo, for example, has been known to rate consumers as less creditworthy if they are friends on Facebook with someone who was late paying back a loan.[48] Algorithmic bias is discussed later in this chapter and in chapters 4 and 6.

Payment systems. Digital payments are by definition flows of electronic data. They are central to powering e-commerce and other online transactions, while simultaneously generating data on purchasing patterns that can provide insights into a plethora of consumer characteristics.

Mobile payments in particular have reduced the transaction costs of transferring resources, enabling new ways for households and firms to make payments, save, and send remittances. The well-documented benefits of mobile money in lower-income countries include lowering transaction and transport costs; encouraging saving through the relatively safe storage of value in a digital format; empowering female users through greater privacy, thereby increasing their bargaining power within families; and allowing more effective risk sharing between households.[49] In India, mobile money has improved the ability of households to share risk, providing welfare benefits of 3–4 percent of income on average.[50] Research also has found a significant link between the use of the mobile financial service M-Pesa and a reduction in poverty among Kenyans, with greater impacts on female-headed households through changes in financial behavior and movement of labor from subsistence farming to secondary jobs and entrepreneurship.[51]

Use of transaction data for product development. Digital payments generate large amounts of data on how people make purchases and transfers, which can be especially important in economies that run largely on cash and among demographic groups that have a small digital footprint. These data can enable firms to see which regions and market segments are expanding, understand user preferences and behavior to target services such as microcredit, and predict fraud and increase security within and between platforms. For example, Mastercard's Tourism Insights service allows the tourism industry to make better investments by leveraging big data to provide information on travelers' preferences.[52] In South Africa, TymeBank offers customers incentives to link their debit cards to their retail loyalty programs, providing access to data on customer spending that are used for product design.

Distributed ledger technology, including blockchain. Blockchain eliminates the need for financial intermediaries, drastically reducing settlement time and making transfers almost instantaneous. The use of digital technology can embed rules into smart contracts, including automated execution of contract. The explicit terms and payments of DLT can simplify complex negotiation and verification processes.[53] DLTs' use of smart contracts in the provision of loans and credit can also improve trust. This is especially important for new and smaller firms that lack the requisite credit histories and collateral.

Despite the promise of blockchain, there are serious challenges to its widespread adoption, including unclear or unfavorable regulatory approaches and lack of user understanding. Adopting blockchain where the technology does not address the underlying issue or consumer needs is also problematic.

Agriculture

Managing production and marketing risks is a key challenge for smallholder farmers and agribusinesses. Remote sensing and geographic information systems, together with data analytics, provide insights into farming operations and propel the development of smart farming, which can help manage production and financial risks. For example, NubeSol, an Indian agtech firm, provides sugarcane growers with a monthly yield map of their plots, with forecasts of yields and recommendations on inputs such as fertilizer based on remote sensing and data analytics.

JD Digits (JDD), a technology firm in China, is adopting AI techniques and big data to provide credit to farmers who raise pigs. Farms install AI-enabled cameras that can recognize pigs' faces, as well as IoT technology to transmit data about the farms' physical conditions. If a pig with feeding abnormalities has been identified, the algorithm can quickly extract information about its growth history and immune status to provide customized feeding care. The IoT system adjusts farm conditions such as humidity, temperature, and lighting based on real-time data on the farm. Using information about farm operations, JDD also carries out credit assessments to provide farmers with loans, which has reduced their nonperforming loan ratio to nearly zero.

Platforms are using data as well to provide a range of services and products along the value chain, including by reducing idle capacity in machinery. Hello Tractor, which emerged in Nigeria, operates a platform connecting tractor owners and farmers who lack their own equipment. Data about tractor locations and availability are monitored using an installed device and then transmitted to Hello Tractor's mobile app platform, which farmers can use to submit a booking request. In this way, farmers are able to find the most cost-effective available tractor, and tractor owners are able to monitor the use of equipment. Another agriculture platform, DigiCow, pioneered in Kenya, keeps digital health records on cows and matches farmers with qualified veterinary services.

Integrated, data-focused solutions are emerging along the whole agriculture value chain. Digifarm, a mobile platform offered by Safaricom in Kenya, provides farmers with one-stop access to a suite of products, including financial and credit services, quality farm products, and customized information on best farming practices. Mobile money data from M-Pesa and data on the way people behave on the app are taken into consideration to provide farmers with tailored products and services.

As agriculture supply chains become more complex, margins imposed by different intermediaries mount, which raises the prices paid by consumers and depresses the income earned by farmers. Food traceability concerns also increase.[54] Data-based solutions can improve food traceability, while disrupting traditional market structures by reducing the need for intermediaries. In Haiti, blockchain solutions have allowed mango farmers to maintain ownership of their produce until the final sale to US retailers by facilitating traceability and direct payments. Employed in conjunction with other value chain components such as third-party logistics services, intermediaries that previously held substantial market power are circumvented. Customers can scan a QR code on the final product to access information about where the mango comes from, how it was packaged and transported, and the costs involved at each step.[55] Similarly, Walmart has collaborated with IBM to trace mangoes from South and Central America to the United States. Participants in this process cannot edit information because of the decentralization feature of blockchain technology, which ensures trust and transparency.[56]

Personal data protection has specific complexities for farmers. Data on their farms are identifiable and could be used to reveal personal details such as their wealth and income. However, farmers could also benefit from using and pooling their data to develop commercial insights. The governance regime for agriculture thus requires special considerations (see the further discussion of governance issues in chapter 8).

Health

To deliver individual health care in lower-income countries, data-driven applications require complementary improvements in infrastructure and basic health services before they can become truly transformational. Some business models show promise in helping overcome such challenges, which include high logistical costs, counterfeiting of pharmaceuticals, difficulties in coordinating health care resources, and low supplies of specialist expertise, especially in rural areas. With strong mobile phone penetration, rising investment in digitizing health information, and developments in cloud computing, more health-focused businesses in low- and middle-income countries are likely to adopt data-intensive advances in coming years. Although such advances hold promise, the sensitive nature of health data implies an acute need for policy makers to be aware of the risks posed by the improper collection and use of these data.

Telehealth (mHealth and eHealth). Telehealth makes use of data and connected devices to deliver care remotely. In rural areas where the ratio of doctors to patients is low, telehealth is a useful way to access consultations and disease diagnosis. The model has also played an important role during the COVID-19 pandemic, where remote diagnosis has been necessary.

Mobile apps combined with AI technology and wearable devices can provide in-the-field diagnoses and recommendations. For example, Colorimetrix, an app that allows a smartphone camera to read results from color-based tests for diabetes, kidney disease, and urinary tract infections, was designed specifically with lower-income economies in mind. Algorithms are used to compare the result with stored calibration values. Results are delivered to the smartphone, allowing for further analysis of results for trends. The hope is that such apps will eventually also be able to detect HIV, malaria, and tuberculosis.[57]

Accuhealth Chile monitors patients in remote areas by using a range of connected medical sensors. Both quantitative data on patients' progress and qualitative data collected through custom-created questionnaires are sent to a virtual clinical service that conducts patient triage based on algorithmic analysis. Accuhealth is also using predictive algorithms to make service delivery more cost-efficient.

In Cameroon, CardioPad was locally designed to improve the access of patients living in rural areas to cardiovascular health care. The CardioPad tablet is paired with sensors that collect data on the patient's health statistics and transmit them over a mobile network to hospitals where cardiologists can make a diagnosis.

Drug verification. Substandard or falsified medical products will be an urgent health care challenge in the next decade, according to the World Health Organization (WHO).[58] An estimated one-tenth of medical products in low- and middle-income countries, particularly antimalarials and antibiotics, are substandard or falsified.[59]

Mobile authentication services such as mPedigree offer people in countries such as Ghana, India, Kenya, and Nigeria an easy way to check the authenticity of medicine. Launched in Ghana in 2007, mPedigree allows pharmaceutical manufacturers to add a code to their packaging that consumers can then verify using their mobile phones. mPedigree has also begun using its consumer authentication data to monitor for anomalies in real time so that it can then generate warnings to brand owners, regulators, and consumers.[60]

Supplies management. Digital platforms can also help manage supplies in countries where centralized provision is deficient or lacking. LifeBank is a Nigerian platform firm that matches hospitals requesting blood with potential donors based on current demand and location maps of all institutions involved in blood distribution. Information about the donation, collection, screening, storage, and delivery procedures are recorded on a blockchain, thereby increasing confidence in blood quality. LifeBank claims to have reduced the average delivery time from about 24 hours to 45 minutes.[61] During the COVID-19 crisis, the platform has also extended its services to matching medical equipment.

Education

Despite significant improvements in school enrollment over the last decades, an average student in low-income countries performs worse than 95 percent of the students in high-income countries.[62] Lack of teaching resources and learning tools and the traditional one-size-fits-all approach in education have made it difficult to tailor instruction to students' individual abilities and needs, particularly where classrooms are overcrowded.[63] Recent advances in big data and AI offer opportunities to provide individualized learning experiences for students. Machine learning and data analytics techniques can help identify students' behavioral patterns (such as mistakes made frequently in tests) usually in a more efficient way than teachers. For example, by memorizing and understanding students' learning paths, Byju, a digital supplemental learning platform in India, suggests tailored learning materials such as videos, quizzes, and flashcards that match the needs of individual students. Besides analyzing individual learning behaviors, the platform also analyzes aggregate data on how all students learn on its platform. If many students are having trouble with similar types of problem sets, the system flags the need to add more explanatory videos or materials to the entire platform.

In China, Squirrel Ai Learning is another firm specializing in intelligent adaptive education. Students start with a short diagnostic test to leave a digital footprint reflecting their knowledge level so that the teaching system can provide a tailored curriculum, which is updated as the student proceeds through learning modules. Based on its comparison trials among middle school students, Squirrel Ai Learning claims that its system does a better job of improving math test scores than traditional classroom teaching.[64]

During the COVID-19 pandemic, digital platforms that support live video communication have been playing an indispensable role in transitioning to online learning. As of May 2020, more than 140 countries had closed schools, affecting more than 60 percent of enrolled students.[65] Lark, for example, is providing educational institutions in India with free cloud storage and video conference services. Dingtalk, a communication platform that supports video conferencing and attendance tracking, has connected more than 50 million students with teachers in China.

Transport and logistics

Data-driven firms in transport and logistics provide matching services to facilitate the use of assets by other market participants. By automating decision-making and navigation, these models increase the efficiency of service delivery and the management of supply chains. Prominent applications are for digital freight matching, digital courier logistics, and IoT-enabled cold storage.

Digital freight matching. These platforms (often dubbed "Uber for trucks") match cargo and shippers with trucks for last-mile transport. In lower-income countries, where the supply of truck drivers is highly fragmented and often informal, sourcing cargo is a challenge, and returning with an empty load contributes to high shipping costs. In China, the empty load rate is 27 percent versus 13 percent in Germany and 10 percent in the United States.[66]

Digital freight matching overcomes these challenges by matching cargo to drivers and trucks that are underutilized. The model also uses data insights to optimize routing and provide truckers with integrated services and working capital. Because a significant share of logistics services in lower-income countries leverage informal suppliers, these technologies also represent an opportunity to formalize services. Examples include Blackbuck (India), Cargo X (Brazil), Full Truck Alliance (China), Kobo360 (Ghana, Kenya, Nigeria, Togo, Uganda), and Lori (Kenya, Nigeria, Rwanda, South Sudan, Tanzania, Uganda). In addition to using data for matching, Blackbuck uses various data to set reliable arrival times, drawing on global positioning system (GPS) data and predictions on the length of driver stops. Lori tracks data on costs and revenues per lane, along with data on asset utilization, to help optimize services. Cargo X charts routes to avoid traffic and reduce the risk of cargo robbery. Kobo360 chooses routes to avoid armed bandits based on real-time information shared by drivers. Many of the firms also allow shippers to track their cargo in real time. Data on driver characteristics and behavior have allowed platforms to offer auxiliary services to address the challenges that truck drivers face. For example, some platforms offer financial products to help drivers pay upfront costs, such as tolls, fuel, and tires, as well as targeted insurance products.

Kobo360 claims that its drivers increase their monthly earnings by 40 percent and that users save an average of about 7 percent in logistics costs.[67] Lori claims that more than 40 percent of grain moving through Kenya to Uganda now moves through its platform, and that the direct costs of moving bulk grain have been reduced by 17 percent in Uganda.[68]

Digital courier logistics. The growth of on-demand couriers enables small merchants and the growing e-commerce industry to reach customers rapidly and reliably in expanding urban areas. Data-driven matching and route optimization overcome high search costs and traffic congestion and provide verification of safety standards through customer reviews.

The prime example of this business model is Gojek, which is reportedly utilized by more than 1 million motorcycle drivers serving 500,000 micro, small, and medium enterprises (MSMEs)—including more than 120,00 MSMEs since the onset of the COVID-19 pandemic.[69] Established in Indonesia in 2010 as a call center to connect consumers to courier delivery services, the company leveraged its data on consumer behavior to expand into digital courier services in 2015. Its app now offers various logistics services, including delivery of food and groceries and medicines and pharmaceuticals. Gojek uses AI and ML for matching, forecasting (to inform drivers where to go ahead of a surge in demand), and dynamic pricing. Through its 8 billion pings with drivers per day, Gojek claims it generates 4–5 terabytes of data every day.

IoT-enabled cold storage. According to WHO, 19.4 million people across the globe lacked access to routine life-saving vaccines in 2018, partly because of lack of efficient cold chain systems.[70] IoT-enabled cold storage solutions allow the transport and storage of temperature-sensitive food and medication, with greater control and tracking by the freight owner. For vaccines, the ability to track temperature can help ensure confidence in integrity before dispensation, even before further testing.

Gricd, a Nigerian start-up founded in 2018, utilizes solar-powered, IoT-enabled mobile refrigeration boxes whose internal probes collect temperature data and transmit it to a server. Real-time information on location and temperature can be accessed by freight owners online or via a mobile app, ensuring that the cold chain is effectively monitored and maintained.

Social media as a tool for connecting to markets

High marketing and advertising costs hinder smaller businesses trying to reach new markets and customers. Meanwhile, high search costs and frictions related to contract enforcement raise prices for buyers and inhibit trade.[71] Social media provide a low-cost sales platform for firms domestically and abroad and reduce search costs for consumers. They enable products to be better targeted to consumers and can reduce marketing costs by as much as 90 percent, compared with traditional television marketing.[72] Social media platforms also allow sellers to incorporate market intelligence into their product development through real-time feedback and gathering of online data.[73] Given these advantages, it is not surprising that nearly half of all enterprises in the European Union had used social media for advertising purposes as of 2017.[74]

Reaching markets through social media could disproportionately advantage smaller firms over larger ones, as suggested by the high proportion of small entrepreneurs who use Facebook. In 2018, nearly four in 10 Facebook business users were single-person firms connecting across 42 countries (including low- and middle-income), although single-person firms are only one-tenth of the general population of firms.[75] Businesses run by women are more likely to leverage online tools to facilitate business success than businesses run by men.[76]

On the buyer side, social connections can increase trade by building trust, including by reducing information asymmetries and providing a substitute for the formal mechanisms of contract enforcement.[77] According to a study of 180 countries and 332 European regions, social connectedness tends to increase exports—particularly to those countries with a weak rule of law—and to lower prices, especially for goods whose prices are not transparent and that are not traded on exchanges.[78]

Some potential risks and adverse outcomes of data-driven businesses to be addressed by policy

Despite the potential transformative effects of data-driven firms, policy makers need to take into account several (often interrelated) risks and adverse outcomes to ensure that the use of data in the productive processes of firms safely fulfills their potential. The relevance and immediacy of these concerns depend on the data intensity of a country's economy. However, because of the global nature of many large data-driven firms market dynamics in one country can often have spillover effects internationally. This concern should not discourage policy makers from fostering a data-driven economic ecosystem in their country, but they should put the appropriate safeguards and enablers in place to ensure that data-driven markets remain competitive and vibrant—and that gains are shared broadly across society—as the data intensity of the economy increases. These topics are covered in part II of this Report.

Potential to increase the propensity for dominant firms to emerge

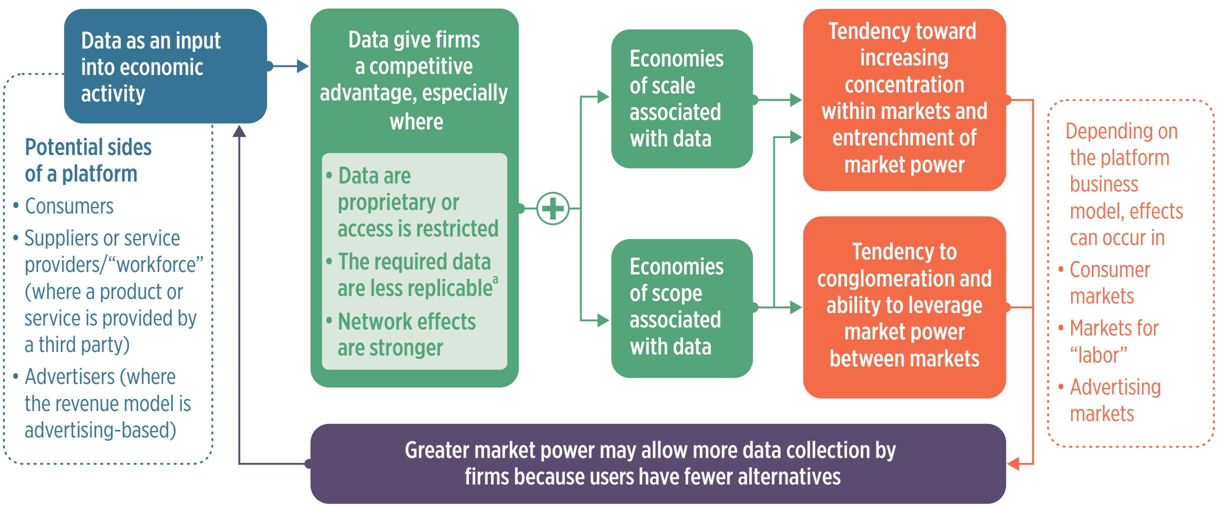
Proprietary data can provide a firm with a competitive advantage over rivals. Because data are often created as a by-product of a firm's economic activities, once a firm has invested in the fixed cost of building capacity to collect data, the marginal cost of creating additional data is low. Moreover, better targeting of a firm's offering can attract more users, thereby leveraging network effects between platform users that can lead to a "winner-takes-most" dynamic or, at the least, a scale advantage that new entrants find difficult to overcome.[79] For example, an e-commerce platform that incorporates more consumer data creates a more customized shopping experience, with more accurate product recommendations, more preordered shopping baskets, and more consumer reviews. A platform with a greater number of consumers will also attract more suppliers through indirect network effects, raising users' costs of switching to competing platforms. The distribution of web traffic, a proxy for concentration in the e-commerce sector, is skewed toward a few larger platforms. Among 631 business-to-consumer online marketplaces in Africa, 56 percent of web visitors went to 1 percent of sites in 2019.[80] Jumia alone had 24 percent of users.

Data can also ease a platform's entry into adjacent markets. Well-known examples are M-Pesa's move from money transfers into savings and loan products; Uber's entry into food and freight delivery; and Google's evolution from search to shopping, maps, and other markets. By combining multiple types of data, platforms can benefit from the broader scope of their data, which has spurred a growing number of mergers aimed at accumulating data (a prime example is Facebook's acquisition of WhatsApp).[81]

The potential of a platform business to acquire market power depends on its business model, including the types of users that interact on the platform and its revenue model, which affect the type of data that gives firms a competitive advantage (figure 3.7). For example, firms that rely on advertising revenue require frequently updated consumer data to create holistic consumer profiles. Platforms that earn a fee based on their transactions conducted may rely more on historical data on product demand and consumer profiles for a smaller range of products. The type of data required affects, in turn, the ease with which firms can access or replicate the data they need. Where platforms rely on volunteered or observed consumer data, firms with greater market power may be able to collect data more easily because consumers have fewer options—meaning those firms can further entrench their market positions.

The greater propensity for dominance in data-driven markets raises the risk that smaller or more traditional firms will be excluded, hindering local entrepreneurship and posing risks for consumer welfare. These effects can be exacerbated in developing markets, where entrants find it harder to raise start-up capital and hire from the limited supply of skilled programmers and data scientists. For example, of the total private market funding received by the 10 highest-funded disruptive tech firms in Africa, 77 percent went to firms owned by the three largest African internet companies (two by Naspers, two by Jumia, and one by Ringier One Africa Media).[82] Where few large data-driven players currently operate or where a few large firms provide much-needed goods and services, the risks may be less immediately apparent. However, because of the dynamism of such markets and their tendency to tip toward concentrated structures, it is important that policy makers safeguard against dominance that forestalls entry and innovation.

Figure 3.7 Risks to market structure and market power stemming from platform firms



Source: WDR 2021 team.

a. Such data include observed and inferred data and data requiring frequent updating.

On DLT platforms, data access is not controlled by one institution, thus reducing the extent to which the benefits of network effects can entrench market power. However, private blockchains still require a central authority or institution that decides who can participate in the system, thereby placing power in the hands of the institution that acts as the gatekeeper. By contrast, public blockchains such as Bitcoin do not require a central authority but rather "proof-of-work" (a system to deter frivolous or malicious uses of computing power) to participate. This system, in turn, generates very high energy costs because it requires computing power and resource usage.[83]

Phenomena linked to the issue of dominance and market structure are described in the sections that follow.

Tension between cooperation and competition in data-driven ecosystems. Complementary products are built around larger platforms. These innovations can become central to the business models of the larger platforms**.** For example, Amazon, Uber, and Airbnb would not be able to operate without the payment systems embedded in their services. However, this network structure also means that firms are dependent on accessing the systems and data of other firms that could become rivals in the future.

Typically, firms access the systems and data of other firms through application programming interfaces (APIs) provided by the core platform in the ecosystem. APIs link platforms to other platforms and to developers of digital services. Through APIs, a platform or digital service provider will typically either draw data from or provide data to other firms to support its own functioning or support the functioning of other players. In markets where data are a key input, the owners of valuable data are gatekeepers in the development of smaller entrepreneurs.

The emergence of potential competition from complementors may provoke the lead platform to restrict its API. For example, when Twitter perceived a competitive threat from LinkedIn, Twitter restricted the use of its API in 2012 to prevent users' Tweets from appearing on LinkedIn's platform. In Kenya, developers have complained about M-Pesa's refusal to share its API.[84] In the United States, developers testifying before Congress in 2020 accused Apple of mimicking their products and of citing privacy concerns to restrict how third-party developers collect location data.[85] Chapter 7 covers competition issues.

Data-driven mergers and acquisitions. In recent years, waves of acquisitions by large local players have occurred in e-commerce in China and India and in transportation in Southeast Asia, where Uber exited eight Southeast Asian markets after selling its businesses to Grab, the region's leading platform.[86] Of the mergers involving digital platforms that have undergone review by antitrust authorities around the world, 82 percent involved an acquisition by a very large firm.[87] Mergers of two very large firms were the most common type of transaction. Chapter 7 covers issues related to mergers in more detail.

Suboptimal exchange of data. Although the broad use, reuse, and repurposing of data by firms can generate larger gains, market mechanisms may generate specific patterns of data exchange and reuse below the level that yields the greatest social welfare.[88] Regulators must take several steps to balance the costs and benefits of mandating data sharing to address these concerns. First, they should protect individuals' rights related to personal data. Second, they should recognize that mandated data sharing would dampen firms' incentives to invest in data collection if firms must share data with competitors or potential competitors. And, third, they should take into account that data sharing could jeopardize the provision of free or subsidized services if a firm relies on monetization of its data to cross-subsidize these services. Optimal data sharing between firms could be lower in countries where data are less important to the economy.[89] The right balance may differ in high-income and lower-income economies, although policy makers in some lower-income countries may wish to adopt a forward-looking viewpoint in this area to set the stage for future advances in their data economies. These issues are further discussed in chapters 6 and 7.

Linked to data sharing is the debate over property rights or access rights to data generated as a by-product of economic activity—for example, in terms of individuals versus firms in the case of personal data or in terms of owners of devices or applications versus the party using the device or application. This issue is taken up in chapter 6.

Potential for exploitation of individuals

Excessive data collection. Data collected by firms tracking users across third-party websites, applications, and devices can raise concerns. This practice is dominated by a small number of large firms—for example, a major part of Google's data collection occurs when a user is not directly engaged with any of its products.[90] Recent literature suggests that the vast amounts of data collected in this manner may be deemed excessive under existing European competition laws, where the focus is on the anticompetitive harm that may occur whether or not data protection rules are infringed.[91] More generally, both data protection authorities and competition authorities have noted that firms often understate and obscure their actual data practices, preventing consumers from making informed choices.[92] Excessive collection of data on children and other vulnerable groups is of particular concern (such as in the education, health, and financial sectors), especially in countries and locales where individuals lack adequate knowledge to protect against these risks. These issues are further discussed in chapters 6 and 7.

Insufficient governance of data held by private firms. Firms choose how much to invest in cybersecurity and data protection, but that investment may fall short of the level that yields the greatest social welfare because firms do not fully internalize the value of privacy and security to the individual user or the need to engender trust in the data economy. Infrastructure service providers may tend to underinvest in cybersecurity because the economic consequences of any breach in data security are largely borne by the clients whose data are compromised. A 2018 report by the US Council of Economic Advisers estimated that malicious cyberactivity cost the US economy between US$57 billion and US$109 billion in 2016, representing between 0.31 and 0.58 percent of that year's GDP.[93] Governments may need to provide incentives or regulations to offset the tendency of firms to underinvest in cybersecurity, including imposing adequate penalties for data breaches.[94]

Pricing, discrimination, and algorithmic risks. Because firms have so much information on customers' personal characteristics and purchase histories, they can adjust their offerings to charge higher or lower prices based on an individual's price sensitivity. This practice can allow low-income sectors of society to be served that otherwise would not be. It can also help firms clear their stock, thereby reducing wastage. However, such price and quality discrimination can also harm some consumers. This kind of price discrimination is not inherently bad—it is a transfer of surplus from consumers to producers, and governments can use taxes and transfers to distribute it back again—but data and data-driven business models do make it easier to discriminate by price.

Algorithms can facilitate anticompetitive behavior by firms, ultimately harming individuals through higher prices. Algorithms can be trained to collude independently by surreptitiously following the behavior of a price leader, or they may be unintentionally biased because of inherent bias in their training data. This bias can magnify marginalization because the most vulnerable populations are often those that are least represented in digital data. Although algorithms hold the promise of impartiality, this promise is not always realized because ultimately they reflect the same biases in human judgment and behavior (due to logic bias and flawed assumptions) reflected in their training datasets. But because algorithmic decision-making is opaque, the potential biases and anticompetitive effects may be difficult to detect.

In algorithmic decision-making, including credit scoring, these risks could lead to discrimination along socioeconomic lines that entrenches existing inequalities.[95] Groups with more limited access to mobile phones, the internet, and bank accounts, such as women, may become less visible in data and decision-making if algorithmic bias is perpetuated through use of biased datasets. If decisions are based on data about those with whom a person interacts, such as friends and neighbors, this, too, may amplify discriminatory effects. For example, a poor credit score for an individual may contribute to lower scores for those in their neighborhood or social network. Furthermore, alternative scoring tools may be used to identify vulnerable individuals susceptible to predatory loans and other product offerings.[96]

These concerns suggest the need to establish a system of oversight, inspection, and auditing of firms' algorithms. However, adequate standardized legal and regulatory frameworks to deal with risks from AI and regulatory capacity to determine harm and the appropriate safeguards are lacking.

Indirect management of the workforce through algorithms. Remotely collecting data on workers and service providers to drive automated or semiautomated decision-making on parameters such as task allocations, performance evaluations, and incentives for certain types of behavior has become particularly prevalent in the gig economy.[97] In addition to the risks algorithmic management raises for bias and discrimination, the practice makes it easier for firms to avoid classifying individuals as employees and thus avoid providing workers with benefits. Better understanding of the organizational and welfare impacts of algorithmic management and data collection on workers would help identify appropriate protections.

Potential to increase inequality within and among countries

Adoption of data-driven business models could widen gaps within countries, between countries, and between different types of firms, different types of workers, and individuals in different income groups.

Within a country, the impact of the data-driven economy on individuals—as consumers, entrepreneurs, or job seekers—will depend on their access to finance, education levels, skills, and technology. Although selling through platforms can close the productivity gaps between large and small firms for those small firms that go down this sales channel, overall smaller firms and entrepreneurs around the world lag their larger counterparts in adopting basic technologies such as fast broadband, having an internet presence, selling online, and utilizing cloud computing. Although the gig economy provides opportunities for job creation, only those who have the assets and skills to participate (such as cars, mobile devices, and literacy) will be able to benefit. And even though automated decision-making may mean more efficient and cost-effective service delivery for some individuals, it may lead to greater bias and discrimination against others.

Likewise, the degree to which a country can benefit from the data-driven economy depends on its underlying infrastructure, capabilities, and scale. The amount of data that can be derived locally depends on a country's level of digital economic activity. Firms from larger, more connected economies—or firms that already operate across countries—with access to larger datasets will have an advantage that only grows with time. Firms from low- and middle-income countries are more likely to lack both access to finance to cover the initial costs of collecting and managing their data and the analytical capabilities to derive value from them. When combined with fewer (or more uncertain) opportunities for monetizing data, either now or in the future (such as through advertising or development of new products), firms from lower-income economies also have less incentive to invest in collecting and analyzing data, which can worsen inequality between countries on a macro level.

Discouraging international data-driven firms from operating or locating in lower-income countries (such as through restrictive data policies) is not a viable solution because it deprives the local economy of the pro-growth and development benefits that data-driven firms can provide. Moreover, it prevents the development of a local ecosystem of data-driven entrepreneurs built around these larger firms—a scenario that could slow the advancement of infrastructure and capabilities needed for lower-income countries to bridge the gap in the longer term. Instead, governments can seek to harness the positive welfare effects of the data-driven economy while mitigating the risks to inequality through a combination of digital inclusion policies, public investments, and robust legal and regulatory tools. These are the topics of part II of this Report.

Notes

1. EIU (2012); Manyika et al. (2011).
2. Fernando (2021).
3. Manyika et al. (2011).
4. Brynjolfsson, Hitt, and Kim (2012); Brynjolfsson and McElheran (2016b).
5. von Mörner (2017).
6. Magalhaes and Roseira (2017); Stott (2014).
7. Evans and Schmalensee (2016); Gawer (2014).
8. This figure is according to the database constructed by the i2i Facility. The database covers eight Sub-Saharan African countries: Ghana, Kenya, Nigeria, Rwanda, South Africa, Tanzania, Uganda, and Zambia. See Africa's Digital Platforms Database, Insight2Impact (i2i), http://access.i2ifacility.org/Digital_platforms/.
9. Evans (2016); Evans and Gawer (2016).
10. The countries in the sample are Armenia, Bangladesh, Brazil, Colombia, the Arab Republic of Egypt, Indonesia, Kenya, Malaysia, Morocco, Nigeria, Peru, the Philippines, the Russian Federation, South Africa, Sri Lanka, Tunisia, and Ukraine. These countries were selected based on data availability and to provide a combination of countries of different sizes and levels of economic development across regions.
11. See Nyman and Stinshoff (forthcoming), who base their data on information provided by Crunchbase, Crunchbase (database), https://www.crunchbase.com/; Dow Jones and Company, Factiva (database), https://professional.dowjones.com/factiva/; Thomson Reuters Foundation, "Inclusive Economies," http://www.trust.org/inclusive-economies/. E-commerce includes both business-to-business (B2B) and business-to-consumer (B2C) business models (excluding agriculture wholesale). Transport includes passenger transport and freight transport/logistics, with the latter accounting for about 67 percent of firms in this category on average. Agriculture includes platforms where the main business focus is agriculture, including both marketplace and financial services. Tourism includes booking platforms and accommodation sharing.
12. Data on the founding year were available for 75 percent of firms in the database.
13. A firm is considered closed if the website cannot be found or accessed via internet research, or if the domain is for sale. A firm is confirmed closed if an article noting the firm's exit was found.
14. Sites are ordered by their Alexa traffic rank, calculated using a combination of average daily visitors and pageviews over the past month. The site with the highest combination of visitors and pageviews is ranked 1. See Alexa Internet, "The Top 500 Sites on the Web, by Country" (accessed April 2020), https://www.alexa.com/topsites/countries.
15. Analysis of data from World Bank, Digital Business Indicators (database), https://www.worldbank.org/en/research/brief/digital-business-indicators.
16. Total session time of over 85 billion hours was recorded from May to June 2018, according to data from 2020 on data software development kits, mobile app downloads,

and revenue and usage data from Apptopia, Data (database), https://apptopia.com/.

17. Kumar and Muhota (2012). See UN Global Working Group on Big Data for Official Statistics for an overview of mobile phone data types and their potential use (UNGWG 2019).
18. Bertrand and Kamenica (2018).
19. Berg et al. (2018).
20. Platforms, IoT devices, and blockchain reduce asymmetric information about the quality and trustworthiness of smaller suppliers. For platforms, user reviews partly serve this purpose. For example, better-rated sellers on eBay have higher prices and higher revenues—see Houser and Wooders (2005); Livingston (2002); Lucking-Reiley et al. (2007); Melnik and Alm (2002)—and sellers with low ratings exit from eBay's platform (Cabral and Hortaçsu 2010).
21. Lendle et al. (2012).
22. Lendle et al. (2012).
23. Luo and Niu (2019).
24. De and Raychaudhuri (2008).
25. Balchin et al. (2016).
26. Data are as of 2018 and come from World Trade Organization, Trade Profiles (database), https://www.wto.org/english/res_e/statis_e/trade_profiles_list_e.htm.
27. Balchin et al. (2016).
28. Borchert and Mattoo (2019).
29. The 10 countries are Belgium, France, Germany, Hungary, Italy, Poland, Spain, Sweden, the United Kingdom, and the United States.
30. Bailin et al. (2019).
31. Falk and Hagsten (2015).
32. Calculations of the European Investment Bank (EIB) and World Bank based on EIB (2019).
33. Aghion, Jones, and Jones (2017); Cockburn, Henderson, and Stern (2019).
34. Brynjolfsson and McElheran (2016a).
35. MGI (2018).
36. IFC (2009).
37. Eisenmeier (2018).
38. Bughin and van Zeebroeck (2017). The authors estimate that tapping latent demand could increase industry size by 0.5 percent a year.
39. Brynjolfsson et al. (2019).
40. Alphabet Inc. (2019); Clement (2020).
41. Alibaba Group (2019); McNair (2018).
42. West and Biddle (2017).
43. Singh (2020).
44. Demirgüç-Kunt et al. (2018).
45. Manyika et al. (2016).
46. Ant Financial is an affiliate company of the Alibaba Group.
47. LenddoEFL (2020).
48. Lobosco (2013).
49. Aron and Muellbauer (2019).
50. Patnam and Yao (2020).
51. Suri and Jack (2016).
52. Mastercard (2017). In May 2020, Mastercard and the Caribbean Hotel and Tourism Association (CHTA) launched a Tourism Insights platform that looks at travel trends from search patterns to in-market spending for the Dominican Republic, Jamaica, and Puerto Rico (*Jamaica Observer* 2020).
53. Baruri (2016).
54. Creydt and Fischer (2019).
55. Open Access Government (2019).
56. Kamath (2018).
57. Levy (2014).
58. The World Health Organization (WHO 2018) defines as substandard "authorized medical products that fail to meet either their quality standards or specifications, or both" and falsified "medical products that deliberately/fraudulently misrepresent their identity, composition or source."
59. WHO (2018).
60. Taylor (2016).
61. Google (2021).
62. World Bank (2017).
63. Rouhiainen (2019).
64. Hao (2019).
65. UNESCO (2020).
66. Future Hub (2020).
67. Gerretsen (2020).
68. Okello (2018).
69. Estimates are provided by Gojek. See Universitas Indonesia (2020).
70. WHO (2019).
71. See, for example, Aker (2010); Allen (2014); Eaton and Kortum (2002); Jensen (2007); Simonovska and Waugh (2014); Startz (2017).
72. See LYFE Marketing "Traditional Media vs. Social Media Advertising: Cost Comparison," https://www.lyfemarketing.com/traditional-media-versus-social-media/.
73. Rumo Arongo Ndiege (2019).
74. EU Open Data Portal, Eurostat, "Social Media Use by Type, Internet Advertising" (dataset), https://data.europa.eu/euodp/en/data/dataset/MTxwCIIEx8RhOhZMmgWvg.
75. Facebook, OECD, and World Bank (2017).
76. Facebook, OECD, and World Bank (2017).
77. Bailey et al. (2020).
78. Bailey et al. (2018).
79. A model developed by Farboodi et al. (2019) shows that data accumulation increases the skewness of firm size distribution as large firms generate more data and invest more in active experimentation. Although there has been less research on DLT applications, Benos, Garratt, and Gurrola-Perez (2019) suggested that similar dynamics would also push these markets to concentration.
80. See ecomConnect, International Trade Centre, African Marketplace Explorer (dashboard), https://ecomconnect.org/page/african-marketplace-explorer.
81. Argentesi et al. (2019).
82. CB Insights (2020). Private market funding refers to the total amount of money a firm has received, including from financial institutions and venture funding. Money raised in public markets is excluded.
83. IRGC (2017).
84. Riley and Kulathunga (2017).
85. Romm (2020).

86. Evans (2016).
87. Very large firms have more than 10,000 employees and more than US$1 billion in revenues. See Nyman and Barajas (forthcoming).
88. In the presence of privacy concerns and negative externalities in data sharing, the market may instead generate too much data sharing. See Acemoglu et al. (2019).
89. Jones and Tonetti (2019).
90. Schmidt (2018).
91. Robertson (2020).
92. Kemp (2019).
93. CEA (2018).
94. Gordon et al. (2015); Kashyap and Wetherilt (2019).
95. McGregor, Murray, and Ng (2019).
96. Hurley and Adebayo (2017).
97. Mateescu and Nguyen (2019).

References

Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar. 2019. "Too Much Data: Prices and Inefficiencies in Data Markets." NBER Working Paper 26296, National Bureau of Economic Research, Cambridge, MA. https://doi.org/10.3386/w26296.

Aghion, Philippe, Benjamin F. Jones, and Charles I. Jones. 2017. "Artificial Intelligence and Economic Growth." NBER Working Paper 23928, National Bureau of Economic Research, Cambridge, MA.

Aker, Jenny C. 2010. "Information from Markets Near and Far: Mobile Phones and Agricultural Markets in Niger." *American Economic Journal: Applied Economics* 2 (3): 46–59. https://doi.org/10.1257/app.2.3.46.

Alibaba Group. 2019. "Alibaba Group Announces March Quarter and Full Fiscal Year 2019 Results." Press release, May 15, 2019, Hangzhou, China. https://www.alibabagroup.com/en/news/press_pdf/p190515.pdf.

Allen, Treb. 2014. "Information Frictions in Trade." *Econometrica* 82 (6): 2041–83.

Alphabet Inc. 2019. "Form 10-K: Annual Report Pursuant to Section 13 or 15(D) of the Securities Exchange Act of 1934." US Securities and Exchange Commission, Washington, DC. https://abc.xyz/investor/static/pdf/20200204_alphabet_10K.pdf?cache=cdd6dbf.

Argentesi, Elena, Paolo Buccirossi, Emilio Calvano, Tomaso Duso, Alessia Marrazzo, and Salvatore Nava. 2019. "Ex-post Assessment of Merger Control Decisions in Digital Markets." With contributions by Elena Salomone and Anna Violini, Lear, Rome. https://www.learlab.com/wp-content/uploads/2019/06/CMA_past_digital_mergers_GOV.UK_version-1.pdf.

Aron, Janine, and John Muellbauer. 2019. "The Economics of Mobile Money: Harnessing the Transformative Power of Technology to Benefit the Global Poor." *VoxEU.Org*, May 7, 2019. https://voxeu.org/article/economics-mobile-money.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2018. "Social Connectedness: Measurement, Determinants, and Effects." *Journal of Economic Perspectives* 32 (3): 259–80.

Bailey, Michael, Abhinav Gupta, Sebastian Hillenbrand, Theresa Kuchler, Robert J. Richmond, and Johannes Stroebel. 2020. "International Trade and Social Connectedness." NBER Working Paper 26960, National Bureau of Economic Research, Cambridge, MA. https://doi.org/10.3386/w26960.

Bailin, Alberto, Peter Gal, Valentine Millot, and Stéphane Sorbe. 2019. "Like It or Not? The Impact of Online Platforms on the Productivity of Incumbent Service Providers." OECD Economics Department Working Paper 1548, Organisation for Economic Co-operation and Development, Paris. https://doi.org/10.1787/080a17ce-en.

Balchin, Neil, Bernard Hoekman, Hope Martin, Maximiliano Mendez-Parra, Phyllis Papadavid, David Primack, and Dirk Willem te Velde. 2016. "Trade in Services and Economic Transformation." SET Report, Supporting Economic Transformation, Overseas Development Institute, London. https://set.odi.org/wp-content/uploads/2016/11/SET-Trade-in-Services-and-Economic-Transformation_Final-Nov2016.pdf.

Baruri, Pani. 2016. "Blockchain Powered Financial Inclusion." PowerPoint presentation, Cognizant, Teaneck, NJ. http://pubdocs.worldbank.org/en/710961476811913780/Session-5C-Pani-Baruri-Blockchain-Financial-Inclusion-Pani.pdf.

Benos, Evangelos, Rodney Garratt, and Pedro Gurrola-Perez. 2019. "The Economics of Distributed Ledger Technology for Securities Settlement." *Ledger* 4 (November): 121–56. https://doi.org/10.5195/ledger.2019.144.

Berg, Tobias, Valentin Burg, Ana Gombović, and Manju Puri. 2018. "On the Rise of the FinTechs: Credit Scoring Using Digital Footprints." FDIC CFR Working Paper 2018-04, Center for Financial Research, Federal Deposit Insurance Corporation, Arlington, VA. https://www.fdic.gov/analysis/cfr/2018/wp2018/cfr-wp2018-04.pdf.

Bertrand, Marianne, and Emir Kamenica. 2018. "Coming Apart? Cultural Distances in the United States over Time." NBER Working Paper 24771, National Bureau of Economic Research, Cambridge, MA. https://doi.org/10.3386/w24771.

Borchert, Ingo, and Aaditya Mattoo. 2009. "The Crisis-Resilience of Services Trade." *Service Industries Journal* 30 (13): 2115–36.

Brynjolfsson, Erik, Avinash Collis, Walter Erwin Diewert, Kevin J. Fox, and Felix Eggers. 2019. "GDP-B: Accounting for the Value of New and Free Goods in the Digital Economy." NBER Working Paper 25695, National Bureau of Economic Research, Cambridge, MA.

Brynjolfsson, Erik, Lorin M. Hitt, and Heekyung Hellen Kim. 2012. "Strength in Numbers: How Does Data-Driven Decision-Making Affect Firm Performance?" In *International Conference on Information Systems 2011 (ICIS 2011)*, vol. 1, edited by Association for Information Systems, 541–58. Red Hook, NY: Curran Associates.

Brynjolfsson, Erik, and Kristina McElheran. 2016a. "Data in Action: Data-Driven Decision Making in U.S. Manufacturing." CES Working Paper 16-06, Center for Economic Studies, US Census Bureau, Washington, DC. https://www2.census.gov/ces/wp/2016/CES-WP-16-06.pdf.

Brynjolfsson, Erik, and Kristina McElheran. 2016b. "The Rapid Adoption of Data-Driven Decision-Making."

American Economic Review 106 (5): 133–39. https://doi.org/10.1257/aer.p20161016.

Bughin, Jacques, and Nicolas van Zeebroeck. 2017. "The Best Response to Digital Disruption." *MIT Sloan Management Review* 58 (4): 80–86. https://sloanreview.mit.edu/article/the-right-response-to-digital-disruption/.

Cabral, Luís, and Ali Hortaçsu. 2010. "The Dynamics of Seller Reputation: Evidence from eBay." *Journal of Industrial Economics* 58 (1): 54–78.

CB Insights. 2020. "The Fintech 250: The Top Fintech Companies of 2020." Research Report, CB Insights, New York. https://www.cbinsights.com/research/report/fintech-250-startups-most-promising/.

CEA (Council of Economic Advisers). 2018. "The Cost of Malicious Cyber Activity to the U.S. Economy." CEA, White House, Washington, DC. https://www.whitehouse.gov/wp-content/uploads/2018/02/The-Cost-of-Malicious-Cyber-Activity-to-the-U.S.-Economy.pdf.

Clement, Jessica. 2020. "Google: Annual Advertising Revenue 2001–2019." Statista, New York. https://www.statista.com/statistics/266249/advertising-revenue-of-google/.

Cockburn, Iain M., Rebecca Henderson, and Scott Stern. 2019. "The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis." In *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay Agrawal, Joshua Gans, and Avi Goldfarb, 115–46. Cambridge, MA: National Bureau of Economic Research; Chicago: University of Chicago Press. https://www.nber.org/books-and-chapters/economics-artificial-intelligence-agenda/impact-artificial-intelligence-innovation-exploratory-analysis.

Creydt, Marina, and Markus Fischer. 2019. "Blockchain and More: Algorithm Driven Food Traceability." *Food Control* 105 (November): 45–51. https://doi.org/10.1016/j.foodcont.2019.05.019.

De, Prabir, and Ajitava Raychaudhuri. 2008. "Is India's Services Trade Pro-Poor? A Simultaneous Approach." Markhub Working Paper 16, Macao Regional Knowledge Hub, United Nations Economic and Social Commission for Asia and the Pacific, Macao SAR, China.

Demirgüç-Kunt, Asli, Leora Klapper, Dorothe Singer, Saniya Ansar, and Jake Hess. 2018. *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*. Washington, DC: World Bank. https://doi.org/10.1596/978-1-4648-1259-0.

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography, and Trade." *Econometrica* 70 (5): 1741–79. https://doi.org/10.1111/1468-0262.00352.

EIB (European Investment Bank). 2019. "EIB Group Survey on Investment and Investment Finance 2019: EU Overview." Economics Department, European Investment Bank, Luxembourg. https://www.eib.org/en/publications/econ-eibis-2019-eu.

Eisenmeier, Siegfried R. J. 2018. "Case Study: Ride-Sharing Platforms in Developing Countries: Effects and Implications in Mexico City." P4P Commission Background Paper 3, Pathways for Prosperity Commission on Technology and Inclusive Development, Blavatnik School of Government, University of Oxford, Oxford, UK. https://pathwayscommission.bsg.ox.ac.uk/sites/default/files/2019-09/ride-sharing_platforms_in_developing_countries.pdf.

EIU (Economist Intelligence Unit). 2012. "The Deciding Factor: Big Data and Decision-Making." Business Analytics, Capgemini, Paris.

Evans, David S., and Richard Schmalensee. 2016. *Matchmakers: The New Economics of Multisided Platforms*. Boston: Harvard Business Review Press.

Evans, Peter C. 2016. "The Rise of Asian Platforms." Emerging Platform Economy Series 3, Center for Global Enterprise, New York. https://www.thecge.net/web/viewer.html?file=/app/uploads/2016/11/FINALAsianPlatformPaper.pdf.

Evans, Peter C., and Annabell Gawer. 2016. "The Rise of the Platform Enterprise: A Global Survey." Emerging Platform Economy Series 1, Center for Global Enterprise, New York. https://www.thecge.net/app/uploads/2016/01/PDF-WEB-Platform-Survey_01_12.pdf.

Facebook, OECD (Organisation for Economic Co-operation and Development), and World Bank. 2017. "Future of Business Survey: Gender Management in Business." Factworks, San Mateo, CA. https://fbnewsroomes.files.wordpress.com/2017/02/future-of-business-survey-gender-management-in-business-january-20171.pdf.

Falk, Martin, and Eva Hagsten. 2015. "E-Commerce Trends and Impacts across Europe." *International Journal of Production Economics* 170 (December): 357–69. https://doi.org/10.1016/j.ijpe.2015.10.003.

Farboodi, Maryam, Roxana Mihet, Thomas Philippon, and Laura Veldkamp. 2019. "Big Data and Firm Dynamics." NBER Working Paper 25515, National Bureau of Economic Research, Cambridge, MA. https://doi.org/10.3386/w25515.

Fernando, Jason. 2021. "Factors of Production." *Investopedia*, January 3, 2021. https://www.investopedia.com/terms/f/factors-production.asp.

Future Hub. 2020. "E-Logistics Transforming Mid-Mile Freight Logistics in Emerging Markets: Stories Told and Untold." *Future Insight* (blog), June 30, 2020. http://www.fhub.io/blog/e-logistics-transforming-mid-mile-freight-logistics-in-emerging-markets.

Gawer, Annabell. 2014. "Bridging Differing Perspectives on Technological Platforms: Toward an Integrative Framework." *Research Policy* 43 (7): 1239–49.

Gerretsen, Isabelle. 2020. "Trucking App Kobo360 Wants to Speed Up Deliveries across Africa." *CNN Business: Innovate Africa*, April 16, 2020. https://www.cnn.com/2020/04/16/tech/kobo-360-trucks-spc-intl/index.html.

Google. 2021. "How Doctors, Dispatch Drivers, and Blood Donors Are Coming Together to Save Lives across Africa." *Search On*, Lifebank, Google, Mountain View, CA. https://about.google/stories/lifebank/#:~:text=By%20designing%20a%20system%20to,take%20your%20time%20to%20solve.

Gordon, Lawrence A., Martin P. Loeb, William Lucyshyn, and Lei Zhou. 2015. "Increasing Cybersecurity Investments in Private Sector Firms." *Journal of Cybersecurity* 1 (1): 3–17. https://doi.org/10.1093/cybsec/tyv011.

Hao, Karen. 2019. "China Has Started a Grand Experiment in AI Education: It Could Reshape How the World Learns." *MIT Technology Review*. https://www.technologyreview.com/2019/08/02/131198/china-squirrel-has-started-a-grand-experiment-in-ai-education-it-could-reshape-how-the/.

Houser, Daniel, and John Wooders. 2005. "Hard and Soft Closes: A Field Experiment on Auction Closing Rules." In *Experimental Business Research: Economic and Managerial Perspectives*, vol. 2, edited by Amnon Rapoport and Rami Zwick, 123–31. Dordrecht, the Netherlands: Springer.

Hurley, Mikella, and Julius Adebayo. 2017. "Credit Scoring in the Era of Big Data." *Yale Journal of Law and Technology* 18 (1): 148–216.

IFC (International Finance Corporation). 2009. "M-Money Channel Distribution Case, Kenya: Safaricom M-PESA." IFC, Washington, DC. https://www.ifc.org/wps/wcm/connect/e0d2a9bd-16b9-4a36-8498-0b2650b9af8b/Tool%2B6.7.%2BCase%2BStudy%2B-%2BM-PESA%2BKenya%2B.pdf?MOD=AJPERES&CVID=jkCVy-n.

IRGC (International Risk Governance Center). 2017. "Governing Risks and Benefits of Distributed Ledger Technologies." IRGC, Geneva. https://irgc.org/wp-content/uploads/2018/09/IRGC.-2017.-Governing-risks-and-benefits-of-DLTs.-Highlights.pdf.

Jamaica Observer. 2020. "Mastercard, CHTA Launch Tourism Insights Platform." May 8, 2020. http://www.jamaicaobserver.com/business-report/mastercard-chta-launch-tourism-insights-platform_193696?profile=1442.

Jensen, Robert. 2007. "The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector." *Quarterly Journal of Economics* 122 (3): 879–924. http://www.jstor.org/stable/25098864.

Jones, Charles I., and Christopher Tonetti. 2019. "Nonrivalry and the Economics of Data." NBER Working Paper 26260, National Bureau of Economic Research, Cambridge, MA. https://ideas.repec.org/p/nbr/nberwo/26260.html.

Kamath, Reshma. 2018. "Food Traceability on Blockchain: Walmart's Pork and Mango Pilots with IBM." *JBBA, Journal of the British Blockchain Association* 1 (1): 3712. https://doi.org/10.31585/jbba-1-1-(10)2018.

Kashyap, Anil K., and Anne Wetherilt. 2019. "Some Principles for Regulating Cyber Risk." *AEA Papers and Proceedings* 109 (May): 482–87. https://doi.org/10.1257/pandp.20191058.

Kemp, Katharine. 2019. "Concealed Data Practices and Competition Law: Why Privacy Matters." Report UNSWLRS 53, University of New South Wales Law Research Series, UNSW Law, University of New South Wales, Sydney.

Kumar, Kabir, and Kim Muhota. 2012. "Can Digital Footprints Lead to Greater Financial Inclusion?" CGAP Brief, Consultative Group to Assist the Poor, Washington, DC. https://www.cgap.org/sites/default/files/researches/documents/CGAP-Brief-Can-Digital-Footprints-Lead-to-Greater-Financial-Inclusion-Jul-2012.pdf.

LenddoEFL. 2020. "Product Briefs." LenddoEFL, Singapore. https://lenddoefl.com/resources.

Lendle, Andreas, Marcelo Olarreaga, Simon Schropp, and Pierre-Louis Vezina. 2012. "There Goes Gravity: How eBay Reduces Trade Costs." Policy Research Working Paper 6253, World Bank, Washington, DC.

Levy, Stephen. 2014. "Colorimetrix App Turns Smartphones into Lab Test Readers." *MD+DI*, March 19, 2014. https://www.mddionline.com/digital-health/colorimetrix-app-turns-smartphones-lab-test-readers.

Livingston, Jeffrey A. 2002. "How Valuable Is a Good Reputation? A Sample Selection Model of Internet Auctions." *Review of Economics and Statistics* 87 (3): 453–65.

Lobosco, Katie. 2013. "Facebook Friends Could Change Your Credit Score." *CNN Business*, August 27, 2013. https://money.cnn.com/2013/08/26/technology/social/facebook-credit-score/index.html.

Lucking-Reiley, David, Doug Bryan, Naghi Prasad, and Daniel Reeves. 2007. "Pennies from eBay: The Determinants of Price in Online Auctions." *Journal of Industrial Economics* 55 (2): 223–33.

Luo, Xubei, and Chiyu Niu. 2019. "E-Commerce Participation and Household Income Growth in Taobao Villages." Poverty and Equity Global Practice Working Paper 198, World Bank, Washington, DC.

Magalhaes, Gustavo, and Catarina Roseira. 2017. "Open Government Data and the Private Sector: An Empirical View on Business Models and Value Creation." *Government Information Quarterly* 37 (3): 101248. https://doi.org/10.1016/j.giq.2017.08.004.

Manyika, James, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. 2011. "Big Data: The Next Frontier for Innovation, Competition, and Productivity." McKinsey Global Institute, New York.

Manyika, James, Susan Lund, Marc Singer, Olivia White, and Chris Berry. 2016. "Digital Finance for All: Powering Inclusive Growth in Emerging Economies." McKinsey Global Institute, New York. https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Employment%20and%20Growth/How%20digital%20finance%20could%20boost%20growth%20in%20emerging%20economies/MG-Digital-Finance-For-All-Full-report-September-2016.ashx.

Mastercard. 2017. "Leveraging Big Data to Drive Tourism Revenue." Mastercard Tourism Insights, Mastercard, Purchase, NY. https://www.mastercard.us/content/dam/public/mastercardcom/na/us/en/documents/tourism-insights-summary.pdf.

Mateescu, Alexandra, and Aiha Nguyen. 2019. "Explainer: Algorithmic Management in the Workplace." February, Data & Society Research Institute, New York. https://datasociety.net/wp-content/uploads/2019/02/DS_Algorithmic_Management_Explainer.pdf.

McGregor, Lorna, Daragh Murray, and Vivian Ng. 2019. "International Human Rights Law as a Framework for Algorithmic Accountability." *International and Comparative Law Quarterly* 68 (2): 309–43.

McNair, Corey. 2018. "Global Ad Spending Update: Alibaba, Facebook, and Google to Capture Over 60% of Digital Ad Dollars in 2019." *eMarketer*, November 20, 2018, Insider Intelligence, New York. https://www.emarketer.com/content/global-ad-spending-update.

Melnik, Mikhail I., and James Alm. 2002. "Does a Seller's Ecommerce Reputation Matter? Evidence from eBay Auctions." *Journal of Industrial Economics* 50 (3): 337–49.

MGI (McKinsey Global Institute). 2018. "Notes from the AI Frontier: Modeling the Impact of AI on the World Economy." Discussion Paper, MGI, New York. https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI

-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.ashx.

Nyman, Sara, and Rodrigo Barajas. Forthcoming. "Antitrust in the Digital Economy: A Global Perspective." World Bank, Washington, DC.

Nyman, Sara, and Clara Stinshoff. Forthcoming. "A Mapping of Digital Platform Firms in Developing Economies." WDR 2021 background paper, World Bank, Washington, DC.

Okello, Ron. 2018. "TechCrunch Battlefield Africa Winner Lori Systems Expands into Three New Countries; Wins Multiple Bids Associated with $4B Railroad Project." *Business Wire*, August 18, 2018. https://www.businesswire.com/news/home/20180918005205/en/TechCrunch-Battlefield-Africa-Winner-Lori-Systems-Expands.

Open Access Government. 2019. "World Bank Blockchain Pilot Shows Fresh Narrative for Haiti's Farmers." *Blockchain News*, March 25, 2019. https://www.openaccessgovernment.org/world-bank-blockchain-haitis-farmers/61205/.

Patnam, Manasa, and Weijia Yao. 2020. "The Real Effects of Mobile Money: Evidence from a Large-Scale Fintech Expansion." IMF Working Paper WP/20/138, International Monetary Fund, Washington, DC. https://www.imf.org/en/Publications/WP/Issues/2020/07/24/The-Real-Effects-of-Mobile-Money-Evidence-from-a-Large-Scale-Fintech-Expansion-49549.

Riley, Thyra A., and Anoma Kulathunga. 2017. *Bringing E-money to the Poor: Successes and Failures*." Directions in Development: Finance Series. Washington, DC: World Bank.

Robertson, Viktoria H. S. E. 2020. "Excessive Data Collection: Privacy Considerations and Abuse of Dominance in the Era of Big Data." *Common Market Law Review* 57 (1): 161–90.

Romm, Tony. 2020. "Companies Burned by Big Tech Plead for Congress to Regulate Apple, Amazon, Facebook, and Google." *Washington Post*, January 18, 2020.

Rouhiainen, Lasse. 2019. "How AI and Data Could Personalize Higher Education." *Harvard Business Review*, October 14. https://hbr.org/2019/10/how-ai-and-data-could-personalize-higher-education.

Rumo Arongo Ndiege, Joshua. 2019. "Social Media Technology for the Strategic Positioning of Small and Medium-Sized Enterprises: Empirical Evidence from Kenya." *EJISDC, Electronic Journal of Information Systems in Developing Countries* 85 (2): e12069. https://doi.org/10.1002/isd2.12069.

Sandvine. 2019. "The Mobile Internet Phenomena Report, February 2019." Sandvine, Fremont, CA. https://www.sandvine.com/hubfs/downloads/phenomena/2019-mobile-phenomena-report.pdf.

Schmidt, Douglas C. 2018. "Google Data Collection." Digital Content Next, New York. https://digitalcontentnext.org/wp-content/uploads/2018/08/DCN-Google-Data-Collection-Paper.pdf.

Simonovska, Ina, and Michael Waugh. 2014. "The Elasticity of Trade: Estimates and Evidence." *Journal of International Economics* 92 (1): 34–50.

Singh, Manish. 2020. "Google Ends Its Free Wi-Fi Program Station." *TechCrunch*, February 17, 2020. https://social.techcrunch.com/2020/02/17/google-ends-its-free-wi-fi-program-station/.

Startz, Meredith. 2017. "The Value of Face-to-Face: Search and Contracting Problems in Nigerian Trade." *VoxDev*, July 31, 2017. https://voxdev.org/topic/firms-trade/value-face-face-search-and-contracting-problems-nigerian-trade#:~:text=My%20estimates%20suggest%20that%20search,one%20sixth%20of%20consumer%20spending.

Stott, Andrew. 2014. "Open Data for Economic Growth." Working Paper 89606, World Bank, Washington, DC. http://documents.worldbank.org/curated/en/131621468154792082/Open-data-for-economic-growth.

Suri, Tavneet, and William Jack. 2016. "The Long-Run Poverty and Gender Impacts of Mobile Money." *Science* 354 (6317): 1288–92. https://science.sciencemag.org/content/354/6317/1288.

Taylor, Phil. 2016. "Marie Stopes Turns to MPedigree for Tracking Technology." Securing Industry, Bromsgrove, UK. https://www.securingindustry.com/pharmaceuticals/marie-stopes-turns-to-mpedigree-for-tracking-technology/s40/a2723/.

UNESCO (United Nations Educational, Scientific, and Cultural Organization). 2020. "Education: From Disruption to Recovery." UNESCO, Paris. https://en.unesco.org/covid19/educationresponse.

UNGWG (United Nations Global Working Group on Big Data for Official Statistics). 2019. "Handbook on the Use of Mobile Phone Data for Official Statistics." Draft, Statistics Division, Department of Economic and Social Affairs, United Nations, New York. https://unstats.un.org/bigdata/task-teams/mobile-phone/MPD%20Handbook%2020191004.pdf.

Universitas Indonesia. 2020. "Dampak Ekonomi Gojek 2019: dan Peran Ekosistem Gojek di Indonesia Saat Pandemi COVID-19." Demographics Institute, Faculty of Economics and Business, University of Indonesia, Depok, Indonesia. https://drive.google.com/file/d/1RgVRNJaszs36-3z__smhsqBegrufhNWw/view.

von Mörner, Moritz. 2017. "Application of Call Detail Records: Chances and Obstacles." *Transportation Research Procedia* (25): 2233–41. https://doi.org/10.1016/j.trpro.2017.05.429.

West, Sarah, and Ellery Roberts Biddle. 2017. "Facebook's Free Basics Doesn't Connect You to the Global Internet, But It Does Collect Your Data." *Global Voices Advox*, July 27, 2017. https://advox.globalvoices.org/2017/07/27/facebooks-free-basics-doesnt-connect-you-to-the-global-internet-but-it-does-collect-your-data/.

WHO (World Health Organization). 2018. "Substandard and Falsified Medical Products." Fact Sheet, WHO, Geneva. https://www.who.int/news-room/fact-sheets/detail/substandard-and-falsified-medical-products.

WHO (World Health Organization). 2019. "The Global Vaccine Action Plan (2011–2020): Review and Lessons Learned." Report WHO/IVB/19.07, Strategic Advisory Group of Experts on Immunization, WHO, Geneva. http://www.who.int/immunization/global_vaccine_action_plan/en/.

World Bank. 2017. *World Development Report 2017: Governance and the Law*. Washington, DC: World Bank.