

# Data Concentration and Competition in Digital Platforms

Hyunsoo KIM Research Fellow, Trade Agreement Team, Department of Int'l Trade, Investment, and Economic Security (kimhs@kiep.go.kr)

## I. Introduction

The growth of the digital economy is being driven by the innovation and diffusion of digital technologies, which has sparked a global digital transformation. In particular, digital platforms are leading growth in the digital economy. Recent regulatory proposals regarding competition in the digital market emphasize the importance of ensuring fair competition to sustain innovation and avoid long-term monopolies. This is due to growing concerns over increasing concentration in the market, as a small number of data-rich companies secure a prominent position across horizontally or vertically connected markets, amassing large user bases. For example, in 2022, the EU enacted the Digital Markets Act (DMA) and Digital Services Act (DSA), highlighting how many online businesses are struggling due to various structural issues related to competition, fairness, and market access in the platform economy.

Policymakers have also pointed out the multi-sided nature of digital platforms, network effects, and the tipping effect caused by data concentration, frequently citing data concentration as a barrier that hinders competition in the digital platform market. In a report on the EU's data strategy, the European Commission mentioned that a small number of global digital platform companies hold a significant portion of the world's data, concluding that this could reduce the incentives for new data-driven businesses to emerge, grow, and innovate. In this context, it has been suggested that increasing the openness of the data held by global digital platform companies could help lower entry barriers to the digital market.

This report aims to identify considerations for introducing policies that mitigate the market dominance of data-rich global digital platforms and protect the openness of the digital ecosystem for potential new entrants. To this end, we will examine the impact of data on

competition in digital platforms. By analyzing how digital platforms operate, acquire user data, and use this as an input for innovation to improve service quality, the report will explore the implications for market competition and derive relevant policy recommendations.

## II. The Impact of Data on Competition in Digital Platforms

Data is utilized to improve the quality of services provided by digital platforms and to expand their user base. To achieve this, digital platforms collect, process, and analyze individual-level data generated during platform usage after obtaining users' consent. For example, open marketplaces like Amazon use not only basic information in user profiles, such as age, gender, and location, but also data collected during product searches, to process and analyze in order to improve search and recommendation algorithms. As more data accumulates, search and recommendation results become more tailored to users, which in turn increases the utility or profit that both users and sellers gain from the platform. Additionally, digital platforms can create new business opportunities by collecting and utilizing data. By collaborating with other digital platforms or entering complementary service markets, they can combine their data with that of other platforms to generate additional commercial value. Simply collecting and storing data may not create significant commercial value on its own; the ability to process and analyze the data is essential to extract insights from it.

Digital platforms have the ability to collect individual-level data, which allows them to analyze and more accurately identify user characteristics. Based on users' expressed interests, platforms can understand their preferences and selectively or strategically offer personalized services tailored to each user. As digital devices become more widely available, making it easier to collect individual-level data, and with the advancement of high-performance computing systems and AI algorithms, the scope of data utilization continues to expand. Although it is difficult to predict the extent to which personalized services will eventually develop or in what forms they may be provided, such services are already in use. Examples of personalized services include Google's targeted advertising and Amazon's product recommendations. Through these services, platform users can reduce search costs, while platforms increase the incentive to use their services through effective matching. The possibility of behavior-based price discrimination, which has been mentioned since the early days of digital platforms, cannot be entirely ruled out. Although experimental behavior-based price discrimination strategies have already faced public backlash and produced negative perceptions, as a result of which it has become difficult to observe platforms explicitly using such strategies today, the potential still exists.

When data is utilized in competition within the digital platform ecosystem, this results in different dynamics compared to traditional markets. For example, platforms that collect

and analyze essential data can use that data to gain insights about their competitors and consumer behavior. They can identify which products or services are in higher demand, allowing them to adjust product pricing, product ranges, or service combinations more effectively. At the same time, by restricting competitors' access to the data or analytical results, similar effects can be achieved. A related case is the European Commission's (hereafter, "the EU Commission") investigation into Amazon. In July 2019 and November 2020, the EU Commission launched two investigations into Amazon's alleged abuse of its dominant market position. The investigations focused on: (i) Amazon's use of data, including listings and transaction details of third-party sellers, for its own retail operations, and (ii) Amazon's preferential treatment of sellers that use its logistics and delivery services, particularly with regard to Buy Box selection and the right to sell under the Prime program. The first case is directly related to the platform's use of data. Amazon uses large-scale data for its retail operations to make decisions regarding the composition of product categories, entry or exit of new sellers, pricing strategies, and inventory management at its distribution centers. Producers wishing to sell on Amazon's platform must enter into a contract that allows Amazon and its affiliates to use the extensive data generated on the platform without additional fees. Through these contracts, Amazon gains ownership of the data and is not restricted from using it commercially. The various detailed information collected from competing retailers

on the Amazon Marketplace – including sales, revenues, pricing, product mix, shipping, cancellations, refunds, and visitor traffic – was used to determine when to launch or discontinue product sales, set pricing policies, manage inventories, and select product suppliers for its own retail operations. There is also a potential risk that even if competitors design new products, services, or business models to offer through the platform, dominant digital platforms can swiftly acquire data on these innovations and imitate them. The ways in which dominant digital platforms with dual roles can leverage data, the impact of this data usage, and its potential are new competition policy issues that have emerged in the digital platform market.

Competition between platforms or within platform ecosystems also occurs through innovation. Innovation allows consumers to access new online marketplaces or services more efficiently and at lower prices, and to quickly engage with a wide variety of digital content. Since such innovation positively contributes to consumer and societal welfare, it is crucial to encourage ongoing innovation in the market. However, innovation in the digital platform market differs from that in traditional markets in several ways. In digital platforms, innovation occurs rapidly and simultaneously on multiple fronts. Innovative services or business models are often experimentally introduced while businesses are operating. Some critics argue that digital platforms expand excessively into complementary service markets, but many of these expansions are experimental,

aimed at fostering innovation. In traditional industries, innovations often happen in research labs, whereas in the digital platform industry, they are tested in real-world scenarios. Furthermore, the role of intellectual property rights in the traditional sense is not as prominent in digital platform innovation. Rather than exclusively leveraging the results of innovation through patents or copyrights, digital platforms focus on bringing innovations to market first in order to secure an initial user base.

Above all, one of the biggest differences between innovation in the digital platform market and in traditional markets is the importance of data in the innovation process. As previously mentioned, companies that secure large amounts of data can utilize this to improve the quality of their products or services. At the same time, they can generate revenue through more accurate targeted advertising, which leads to increased investment capacity and accelerates innovation to further improve their products or services. These processes are sometimes referred to as the user feedback loop or the monetization feedback loop. The user feedback loop is similar to the network effect, as it involves improving the quality of products or services by acquiring more data from a growing user base, which, in turn, attracts even more users. The monetization feedback loop can also amplify its effects within this virtuous cycle, but it differs in that even in situations where the influx of new users is limited, companies can still generate continuous revenue, separate from product or service

sales, allowing them to increase their investment capacity.

At first glance, the feedback loop driven by data in digital platforms seems similar to the so-called “learning by doing” effect in traditional markets. In the learning-by-doing process, the more a company provides goods or services to consumers, the more experience it gains, leading to increased learning and the ability to produce goods or services more efficiently. This is very similar to the user feedback loop in the digital platform market. However, there are several key differences between the feedback loop driven by data and the learning by doing effect. First, learning-by-doing primarily benefits consumers by reducing prices through increased production efficiency, whereas the data-driven feedback loop can deliver greater utility to consumers by improving the quality of products or services, even without price changes. Additionally, the data-driven feedback loop can enhance the quality of products or services while consumers are using them, thereby increasing their utility during consumption. Examples of this include cloud-based services and Over-The-Air (OTA) updates for automotive software. This is a phenomenon not found in the traditional learning-by-doing effect, where consumers enjoy additional utility only in subsequent goods or services after having consumed the previous ones. Due to these characteristics, as mentioned earlier, the pace of innovation can accelerate in the digital platform market. Furthermore, even if a digital platform succeeds in innovation through the feedback loop and gradually gains

market dominance, the degree to which the revenue generated from the feedback loop benefits consumers depends on how monopolistically the platform utilizes the data it acquires, and whether consumers can access or use the data they provide. Since consumers are the primary providers of the key inputs in the innovation process, if they have control over this data, their decisions could influence how the benefits of innovation are distributed between the company and consumers.

The emergence of various complementary services linked to digital platforms and the development of platform ecosystems are expected to have diverse impacts on competition in the digital platform market. Unlike traditional markets, the boundaries of digital platforms and complementary service markets shift rapidly, large user bases are quickly formed and move, and innovation occurs rapidly and simultaneously. As a result, competition in the digital platform market differs from the patterns seen in traditional markets, making it challenging to apply conventional competition policies. In the following section, we will examine the considerations to take into account when applying various measures proposed as regulations for digital platforms, particularly those related to data.

### III. Policy Implications Regarding Data in Digital Platforms

Policy discussions in major countries regarding access to data as a means of promoting competition in the digital platform market are

taking place within the framework of existing legislation. The existing legal framework determines whether access to data is necessary based on whether the data is essential for competition in the relevant market, and whether it can only be obtained from the dominant platform. In competition law, discussions about whether other companies can access a dominant firm's resources have traditionally taken place under the so-called "essential facilities" doctrine. Whether data is essential for competition in a given market cannot be determined uniformly. However, existing literature suggests that narrowly defined essential data does not exist in many markets, and it is also difficult to find such data in markets where large digital platform companies like Google or Amazon mainly operate. This is because businesses can be established and operated with search engines or online marketplaces even without extensive user data. From this perspective, the role of competition law in ensuring access to data under the existing legal framework is inherently limited. These limitations arise not only in relation to the essential facilities doctrine but also in the process of applying competition policy. Competition policy typically takes a long time to implement, while the rapidly changing business environment of the digital platform market requires companies to quickly identify business opportunities in order to enter the market. Even if it is proven that a dominant company has used its data in an anticompetitive manner, it is extremely difficult to impose appropriate remedies that could restore the market to its previous competitive state. Additionally, designing

complex remedies and continuously monitoring compliance with competition policy is not easy, which is why there are ongoing discussions about the need for ex ante regulation. However, ex ante regulation also has its own costs and drawbacks, so the practical benefits of implementing such regulations must be carefully evaluated.

One potential measure to limit the scale of user data held by market incumbents is to restrict the retention period of raw data, such as search queries entered by users or location data. According to Chiou and Tucker (2017), even with a shortened retention period, the learning capacity of platforms inferred from data did not significantly diminish. By applying incremental learning techniques, where new models are trained based on existing models and continuously refined with new data, the platform's learning capacity could remain largely unaffected, even with a reduced data retention period. However, if a significant amount of raw data is quickly deleted, there may be some impact on the flexibility and efficiency of the learning models. This is suggested by the fact that, despite Chiou and Tucker's (2017) findings, many analyzed platforms have extended their data retention periods after initially shortening them, citing the need to enhance personalized services. On the other hand, if data retention periods are shortened, market incumbents with large user bases may still gather significant amounts of data within the limited retention period, while market entrants with smaller user bases may struggle

to collect enough data for meaningful analysis, potentially reinforcing the competitive advantage of incumbents. Additionally, a shorter data retention period could limit the amount of user data shared by incumbents with new market entrants. This would likely restrict the ability of entrants to narrow the gap in insights derived from data compared to incumbents. In summary, explicitly limiting the retention period of user data could have positive implications for privacy protection. It may not substantially sacrifice algorithmic efficiency and could also reduce the risk of data consolidation through mergers, as the incentive for "killer acquisitions" driven by data collection would diminish with shorter retention periods. However, it remains questionable whether shorter data retention periods would enhance the competitiveness of potential entrants and promote market entry. As discussed, such a measure could further strengthen the competitive advantage of incumbents and raise barriers to entry.

Another potential solution to ensure competition in the digital platform market is to limit the combination of user data obtained from various services by dominant platforms. This is aimed at preventing the so-called "domino effect" in the digital platform market. To achieve this, data should be stored in separate databases or data silos within the service from which it was originally collected. Under such measures, both incumbents and new entrants would have the same incentives to enter adjacent markets without any inherent competitive advantage. However, these data silos could

hinder the realization of economies of scope that data provides and limit the efficiency that data combination could bring. For example, integrating and interoperating data between email, calendar, and map services could allow users to create appointments from received emails and easily derive travel routes and estimated arrival times from map services. In other cases, combining data from various services may not create synergy but instead harm consumers. For instance, tracking users across different services for targeted advertising purposes could violate their privacy. There are at least two reasons why such measures could face challenges in terms of monitoring and enforcement. First, there is an inherent information asymmetry between regulators and the companies they oversee. It is generally difficult to detect and prove the origin of consumer data and whether datasets have been combined. A notable example is Facebook's acquisition of WhatsApp in 2014. Facebook repeatedly assured the European Commission that it had no plans to link WhatsApp user profiles with Facebook profiles and that such linking was impossible. However, it was later revealed that Facebook knew at the time of the acquisition that it could link user profiles, and in 2016, it publicly announced the integration of profiles. Although some dataset combinations can be monitored through privacy policies, much of the oversight depends on the information provided by the companies themselves. Second, there is ambiguity between competition policy measures and privacy protection regulations. This was evident in the case of the German Federal Cartel Office's order against Facebook,

where Facebook was ordered to separate data collected from Facebook itself and data collected from other services, such as WhatsApp. Under this ruling, consumers must explicitly consent to the combination of data from different sources. This decision could arguably be seen as the competition authority overstepping its boundaries by regulating privacy-related issues. Considering the limitations of data combination policies and the reduction in potential usability that results from restricting dataset combinations, it may not be effective to impose additional obligations at this time. However, it still remains crucial that market incumbents obtain valid and purposeful consent for each service they provide, and the principle of opt-in consent for data combination – rather than opt-out – must be maintained.

**R**estricting the business activities of digital platforms is a strong form of intervention that fundamentally prevents platforms from collecting and combining more user data. In a broader sense, ordering the separation of vertically or horizontally integrated business groups could also be considered a form of business activity restriction. In the past, restrictions or separations were implemented in several network industries, such as the energy, railroads, banking, and broadcasting sectors. However, not all of these interventions have proved effective, and competition authorities are increasingly steering away from such measures, either ex post or ex ante. Nonetheless, there have been continued calls for busi-

ness activity restrictions in the context of digital platforms. For instance, in 2019, India banned foreign e-commerce platforms from directly selling to consumers, and in 2021, a bill was introduced in the U.S. to prohibit online platforms from using their platforms to sell their own products or services, or from owning or controlling other businesses that use their platforms. Restricting business activities would naturally prevent data combination, resolving issues related to monitoring and enforcing data combination policies. Additionally, it is argued that such restrictions could promote diversity and contribute to non-economic policy goals, such as enhancing system resilience. However, there are also downsides to business activity restrictions. First, when restricting business activities or separating certain lines of business, there are several practical challenges. It is often difficult to clearly define the boundaries between different digital markets or services. Furthermore, business activity restrictions, like restrictions against data combination, prevent the positive spillover effects that information generated in one market or service could have on another market. This is one of the main reasons why digital platforms actively expand into adjacent markets. Moreover, as highlighted in existing literature on business separation, splitting or reducing business activities weakens the efficiencies derived from economies of scale and scope. Specifically, the separation of vertically integrated businesses can lead to the problem of double marginalization, creating additional inefficiencies. Therefore, if business activity restrictions are implemented, they should aim to

minimize these inefficiencies. One potential approach is to apply restrictions to the ancillary services of digital platforms. Examples of ancillary services include Identity and Access Management (IAM) services that offer login options like "Sign in with [Platform] ID" or payment services that allow users to pay through third-party services, such as "Pay with [Platform]." These ancillary services obtain user data in exchange for facilitating access to a platform's main service, which users primarily consume. By using such services, ancillary service platforms can track and collect user activity data in markets where they do not operate. Another example of ancillary services is Google Analytics, a tool that collects and analyzes website visitor data to measure and improve the performance of online businesses. These types of services often obtain and utilize more detailed data than what visitors voluntarily provide to the websites. Through ancillary services, dominant platforms can collect various types of user data from multiple markets where they do not directly compete, while having little incentive to innovate the services they provide in those markets. While these ancillary services may not provide better services to consumers by leveraging the data, they can limit the ability of competing firms to gather data and create a data-driven competitive advantage. However, since ancillary services contribute only limited data-based network effects and are often clearly distinguishable from the platform's core business activities, imposing business activity restrictions on them could be relatively straightforward. **KIEP**