

Regulating data platforms from a value of data approach

Arturo Basaure, Juuso Töyli and Petri Mähönen

Abstract

Purpose – This study aims to investigate the impact of ex-ante regulatory interventions on emerging digital markets related to data sharing and combination practices. Specifically, it evaluates how such interventions influence market contestability by considering data network effects and the economic value of data.

Design/methodology/approach – The research uses agent-based modeling and simulations to analyze the dynamics of value generation and market competition related to the regulatory obligations on data sharing and combination practices.

Findings – Results show that while the promotion of data sharing through data portability and interoperability has a positive impact on the market, restricting data combination may damage value generation or, at best, have no positive impact even when it is imposed only on those platforms with very large market shares. More generally, the results emphasize the role of regulators in enabling the market through interoperability and service multihoming. Data sharing through portability fosters competition, while the usage of complementary data enhances platform value without necessarily harming the market. Service provider multihoming complements these efforts.

Research limitations/implications – Although agent-based modeling and simulations describe the dynamics of data markets and platform competition, they do not provide accurate forecasts of possible market outcomes.

Originality/value – This paper presents a novel approach to understanding the dynamics of data value generation and the effects of related regulatory interventions. In the absence of real-world data, agent-based modeling provides a means to understand the general dynamics of data markets under different regulatory decisions that have yet to be implemented. This analysis is timely given the emergence of regulatory concerns on how to stimulate a competitive digital market and a shift toward ex-ante regulation, such as the regulatory obligations to large gatekeepers set in the Digital Markets Act.

Keywords Ex-ante regulation, Data network effect, Value of data, Data complementarity, Data platforms, Digital markets act

Paper type Research paper

Arturo Basaure is based at the Centre for Wireless Communications, University of Oulu, Oulu, Finland. Juuso Töyli is based at the Department of Marketing and International Business, Turku School of Economics, Turku, Finland and Department of Information and Communications Engineering, Aalto University, Helsinki, Finland. Petri Mähönen is based at the Department of Information and Communications Engineering, Aalto University, Helsinki, Finland.

Received 16 June 2024
Revised 1 August 2024
Accepted 15 August 2024

© Arturo Basaure, Juuso Töyli and Petri Mähönen. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

The first author acknowledges funding from the Research Council of Finland (former Academy of Finland) 6G Flagship Programme (Grant Number 346208).

1. Introduction

Data management and related policies have emerged as major topics of interest alongside the development of Internet applications driven by the digitalization of a wide set of services and societal processes. Their increasing significance has raised regulatory concerns over the data policies required to facilitate a competitive ecosystem in emerging platforms and data markets.

Dominant positions in digital services are often explained by network effects, and more recently, to data network effects. While network effects denote that the utility of a service (to a user) increases with the number of service users, data network effects denote an analogous relationship regarding utility and data. These effects can lead to high barriers to entry and oligopoly or monopoly situations with little competition. Thus, given the widespread nature of both effects, data sharing between service providers and consumers generally increases welfare.

According to [Tirole \(2023\)](#), new regulations are needed for emerging digital markets because current regulations do not adequately address antitrust issues related to data. Owing to

network effects, internet platforms can achieve a winner-takes-all position. For example, a monopoly platform is likely to emerge when switching costs are high, network effects are indirect and strong and demand for differentiation is low (Eisenmann *et al.*, 2006). Besides traditional network effects, the accumulation of data provides a platform with an additional advantage for achieving market power. This advantage arises because data may allow increasing returns to scale in terms of value, and the exclusive ownership of data may create significant entry barriers (Rubinfeld and Gal, 2017). Therefore, some authors (e.g. Gregory *et al.*, 2021) have proposed the existence of an additional data network effect. Furthermore, Hagi and Wright (2023) argue that a firm's competitive advantage depends on its ability to learn from data. In this scenario, Parker *et al.* (2020) point out that a regulatory intervention that facilitates data-sharing mechanisms will benefit all market participants, including the consumer.

Beyond the regulatory concern of data management, the value of data itself is a topic of crucial importance. Classical works on the economic value of data (e.g. Moody and Walsh, 2002) indicate that the value of information increases with its quantity (with diminishing returns) and quality (usage, accuracy and combination with other sources) but decreases with time (it is perishable). However, in practice, the value of historical data is subject to controversy. While the availability of historical data was traditionally deemed an advantage (Schaefer and Sapi, 2020), some authors claim that this advantage is often overestimated (Chiou and Tucker, 2017). In general, the real benefit will depend on how much value a particular firm can obtain from the data.

In this context, antitrust regulators are increasingly moving toward ex-ante regulation in digital markets (Yan and He, 2022). For example, the Digital Markets Act (DMA) aims to introduce additional competition by making digital markets contestable. It pushes for ex-ante regulation defining large firms as gatekeepers and setting them obligations by promoting data sharing and limiting some data combination practices for large platforms. While many of the obligations aim to promote competition, some of them may produce unintended counter effects in specific scenarios. For instance, restrictions on the collection of user data can result in data silos and limit value creation (Krämer and Schnurr, 2022). Even more generally, establishing restrictions for players and markets is inappropriate under uncertainty, such as the case of new emerging digital markets (Bauer, 2022).

To address these tensions, this research aims to investigate the impact of regulatory interventions on emerging digital markets, considering data network effects and the economic value of data. We focus on the implications of data sharing and combination practices to market competition and value creation through agent-based simulations.

Overall, we find that the restriction of data combination practices does not incentivize competition, even when applied to very large platforms. In such a case, promoting data sharing and interoperability-enabled provider multihoming seems to function more efficiently as a competition enabler. Data sharing (i.e. portability) fosters competition, while the use of complementary data enhances platform value without necessarily harming competition. Service provider multihoming complements these efforts.

The rest of the paper is organized as follows. Section 2 performs a literature review on the topic. Section 3 presents the scenarios and the model. Section 4 explains the results. Finally, Section 5 concludes.

2. Background

2.1 Background on ex-ante regulation of data

There is increasing concern that the application of ex-post remedies based on competition law may not be sufficient to maintain a competitive regime in emerging digital markets. However, this well-known view (e.g. Stucke and Grunes, 2016) has also been challenged (e.g. Kennedy, 2017), with some scholars claiming that big data-based companies should not be seen as threats but rather as sources of innovation. Under the risk that ex-post

instruments may be ineffective and untimely, regulations should also incorporate ex-ante mechanisms (Kobayashi and Wright, 2020). Ex-post mechanisms have achieved limited efficiency in digital markets (OECD, 2021). First, competition law enforcement has often been too complex, slow and specific to be useful in digital markets with rapidly growing market dominance. Second, in terms of prices and consumer welfare, ex-post instruments may not be suitable for digital markets due to their digital nature, which is based on a two-sided market logic that differs from more traditional markets. Although antitrust analysis can be adapted to platform dynamics, the market requires more proactive regulation when network effects are large (Jullien and Sand-Zantman, 2021). Consequently, there is growing consensus that ex-ante regulation should at least complement the effort of traditional ex-post instruments.

In this general shift toward ex-ante regulation, the EU is in the process of adopting the DMA. The DMA aims to facilitate competition in digital markets using ex-ante regulation, which imposes a variety of prohibitions and obligations to a set of predefined services of large digital platforms. The DMA defines gatekeepers as large platforms behaving as mediators between service providers and consumers. It defines a set of core platform services to which the defined obligations apply and different criteria in terms of turnover and number of users to be considered gatekeepers.

The DMA imposes obligations regarding the collection, processing, access and combination of data (Baschenhof, 2022). The following obligations promote data access and sharing. A gatekeeper should facilitate data portability by end users (Article 5(9)), in line with the General Data Protection Regulation (GDPR). It may also provide access to engagement data to business users (Article 5(10)) and search data to competing search engines (Article 5(11)). However, the DMA also imposes restrictions on data processing and combination practices. For instance, a gatekeeper should not use non-public data in competition with business users (Article 6 (2)), and it cannot combine personal data from a core platform service with any other service or source within its ecosystem (Article 5(2)).

Besides data obligations, the DMA also imposes technical obligations to increase interoperability and decrease end-user switching costs. For instance, a gatekeeper cannot restrict the ability to switch between software applications (Article 6(4)), and it should allow users to uninstall applications that are not essential for the functioning of the OS (Article 6 (3)). Moreover, a gatekeeper must avoid self-preferencing of its services (Article 6(5)).

Some authors have raised concerns about the impact of such an ex-ante regulation. Yan and He (2022) claim that limiting the usage of complementary data within a conglomerate requires further analysis. Krämer and Schnurr (2022) also analyze regulatory remedies that limit the collection and usage of data from complementary sources, highlighting risks such as data siloing and business restrictions. From another perspective, Cennamo *et al.* (2023) claim that the business model agnostic approach of the DMA to self-preferencing and data sharing may risk value creation. Along this same line, Bostoen (2023) criticizes the emphasis of the DMA on market power rather than on value creation. Finally, Heimburg and Wiesche (2023) argue that the DMA is likely to affect the central mode of operation of digital platforms, including openness and governance, because it requires dominant platforms to offer more modularity and limit the control of different layers of ecosystems.

The GDPR has already introduced some restrictions on data processing. For example, an individual could ask a data controller to restrict the processing of their data (Article 18) and establish a general prohibition for decision-making based solely on automated processing (Article 22(1)). The implementation of the GDPR in the EU has shown that digital markets may suffer from restrictive regulations strengthening the position of large firms, even though GDPR effectively delivered positive outcomes in protecting consumer data (Geradin *et al.*, 2021; Gal and Aviv, 2020). For instance, Jia *et al.* (2021) demonstrated a 34.16% decrease in the number of data-related business-to-consumer venture deals after the rollout of the GDPR in the EU in comparison to other geographical areas. Furthermore, Johnson *et al.* (2023) reported an

increase of 17% in market concentration in the website vendor market after the GDPR implementation. Some authors suggest that the value of the GDPR may be expanded through further transparency on how personalization happens, rather than by restricting personalization (Koskinen *et al.*, 2023; Van Buggenhout *et al.*, 2023).

Interestingly, data combination practices have become a major topic of debate, particularly the resulting obligations from Article 5(2) of the DMA. Specifically, the article defines a general prohibition for data combination except when such practice has the user's consent. User consent is introduced in Article 7 of the GDPR, and its implementation provides an advantage for large platforms as they are better prepared to incur associated compliance costs and can more easily obtain user consent from their different products or services (Gal and Aviv, 2020). Consent fatigue from the user has also been identified in the GDPR implementation (Utz *et al.*, 2019), representing an obstacle to the practical and fair expansion of such a mechanism. A study on user consent of the GDPR and DMA claims that the fair implementation of Article 5(2) requires an additional privacy setting solution, in which the user could opt for different types of data combination activities (Botta and Borges, 2023). Other alternative solutions suggested in the literature include a trust structure with visible data-sharing rules for all stakeholders (Koskinen *et al.*, 2023), a sector-wide agreement on personalization transparency (Van Buggenhout *et al.*, 2023), trusted data intermediaries (Podszun, 2022) and bulk sharing of broad anonymized data complemented with personalized user data portability (Krämer and Schnurr, 2022). More generally, data could be considered as labor, with users actively engaging in data work in exchange for economic benefits (Yan and Hen, 2022).

Behind the regulatory discussion on data practices, the value of data emerges as a crucial topic of interest. Data represent an additional network effect, and as such, its strength may determine the competitive outcome. Valavi *et al.* (2021) argue that the value of data (and thus the strength of the data network effect) can be approximated by the efficiency with which machine learning models attain their results. Along this same line, Hagi and Wright (2023) explain that a firm's competitive advantage depends on its ability to learn from data. In particular, such ability depends on the development of technology, the size of the data set and the number of data sources or data complementarity (Kaplan *et al.*, 2020). If data are shared among different participants, then the advantage is also shared, thus reducing restrictions on competition. Parker *et al.* (2020) claim that a regulatory intervention that facilitates data sharing will benefit all market participants, including consumers, because it makes the market more competitive. In addition, due to the non-rivalrous nature of data, firms lack incentives to share, leading to a suboptimal allocation (Jones and Tonetti, 2020). Thus, from a theoretical perspective, the question remains regarding the mechanism and conditions under which data should be shared to incentivize value creation.

2.2 Entry barriers and essential facilities

A classical view in the literature suggests that data constitute an entry barrier and that a large accumulation of data restricts competition (Tirole, 2023). Some authors likewise argue that data are becoming an essential facility and should thus be accessible to competitors (Abrahamson, 2014). Under the essential facility doctrine, a monopolistic firm must share its facilities – in this case, data – with anyone asking for access (OECD, 1996). This condition applies if the exclusive ownership of data is likely to restrict competition, for example, by providing incumbents an advantage in the use of complementary data and by the fact that data may have large economies of scale (Calvano and Polo, 2021).

Regarding the essential facility doctrine, Lambrecht and Tucker (2015) claim that data resources are rarely unique and inimitable but are often substitutable instead. Accordingly, from a resource-based view, data do not constitute an essential facility. Tucker (2019) also highlights many situations where digitalization has weakened the inimitability and uniqueness properties of data, such as data sharing and portability. Therefore, data are unlikely to become an essential facility, and data alone are not very valuable without the capability to extract value from them.

In a similar direction, [Rubinfeld and Gal \(2017\)](#) intensively analyze entry barriers in data markets, including data interoperability and compatibility. They point out that since data are non-rivalrous, those barriers may not be as high as typically thought. In addition, [Evans and Schmalensee \(2017\)](#) argue that the well-known concept of network effect is used incorrectly to explain why big data is bad for the market. They argue that network effects can work in reverse, when, for instance, consumers leave a platform that has lost its value. Therefore, network effects may not always constitute an entry barrier if consumers are not locked-in into platforms because of switching costs.

2.3 Factors and parameters chosen for the analysis

Given the above discussion, in the following sections, we summarize the factors chosen for the analysis to describe the impact of data on value creation and competition. In this analysis, we use [Bain's \(1956\)](#) definition of entry barriers as anything that allows the incumbent to earn above-average profits without the threat of entry. This definition is broader than alternative definitions, such as the one from [Stigler \(1983\)](#), which emphasizes the asymmetries in costs between the incumbent and the new entrant. In line with the previous literature, we consider network effects and switching costs as the main factors explaining entry barriers and the corresponding consumer and provider behaviors (i.e. device lock-in and service provider multihoming) affecting these factors as the parameters. On the consumer side, a user may get locked-in into a platform, for example, since a connected device is not allowed to change to another platform without incurring additional costs (e.g. buying another device). Our analysis assumes a market in which the consumer accesses services through a connected device, such as a mobile phone or a smartwatch. Such costs are considered to diminish over time as the device loses its value. From a service provider perspective, interoperability enabled by modular architecture allows the simultaneous provision of a service through more than one platform (i.e. multihome). When interoperability is lacking between platforms, the costs of providing a service through multiple platforms are higher and act as a switching cost.

Similarly, the inimitability and uniqueness of data define data as an essential facility, and data portability and usage of complementary data are two main policies that may affect these properties. A higher differentiation strategy practiced by a data-driven firm (based on any innovation on extracting value from data) may also increase data inimitability and uniqueness. Even though an essential facility may be considered a type of entry barrier, for this analysis, we separate the essential facility aspects from those making data an entry barrier in a broader sense (i.e. Bain's definition).

The following [Table 1](#) summarizes the factors and parameters chosen for the analysis and the reasons for their selection.

3. Methods

This section presents the scenarios for analysis and the agent-based model for performing the related simulations. In analyzing the defined scenarios, this research uses agent-based modeling and simulation to describe the interaction between different stakeholders in a platform ecosystem under different market conditions and regulatory schemas ([Macal and North, 2014](#); [Tsfatsion and Judd, 2006](#)). This modeling technique has been identified as a

Table 1 Parameters chosen for the analysis

<i>Data as an</i>	<i>Due to (Factors)</i>	<i>Parameters</i>
<i>Entry barrier</i>	<i>Network effect, switching costs</i>	<i>Device lock-in, service provider multihoming</i>
<i>Essential facility</i>	<i>Inimitability and uniqueness</i>	<i>Differentiation, data portability, data complementarity</i>

Source: Created by the authors'

helpful tool for analyzing the suitability of market design and regulation (Bauer and Bohlin, 2022). Agent-based modeling performs a bottom-up study of complex adaptive systems and emphasizes the adaptation of individuals to their environment given a simple set of rules. The present study defines consumers, platforms and service providers as the main interacting roles (i.e. agents) formalized in a horizontally differentiated simulation area. Horizontal differentiation refers to those differences between products and services that do not affect price and quality. These differences interact due to different switching costs and market conditions. Recent examples of agent-based modeling analyzing the dynamics of the ICT ecosystem can be found in several areas, such as applications of the Internet of Things (Basaure *et al.*, 2020), intermediary dynamics of platforms (Köbel and Kunz, 2020), mobile network access dynamics (Finley and Basaure, 2018), two-sided pricing (Sanchez-Cartas, 2018) and platform competition (Huotari *et al.*, 2017).

3.1 Agent-based model

The agent-based model distinguishes two types of service providers, namely, main service providers and complementary service providers, to describe a multi-sided platform, given that emerging data platforms are often intended for multiple types of users. Thus, a consumer values a service provider more if a complementary service provider is also available. For example, a health service provider or a car maintenance service is more valuable for the consumer if an insurance service is also available, and both the provider and the consumer can benefit from each other with higher availability of data.

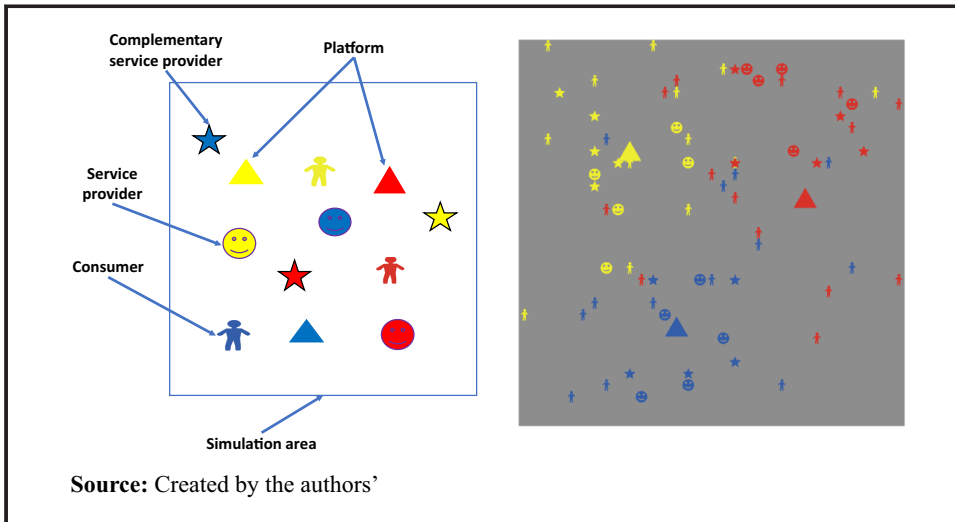
The proposed model analyzes consumers and service providers by locating them in a simulation area that is horizontally differentiated. Horizontal differentiation is based on the Hotelling model of spatial competition (Hotelling, 1929), which describes how firms compete with differentiated products. In its basic form, the Hotelling model describes firms competing in a one-dimensional characteristic space. It uses a cost parameter associated with the distance between products to describe the level of differentiation and identify the best strategies and equilibria on how firms make decisions related to locations and prices. Several authors (Eaton and Lipsey, 1975; Economides, 1986; Veendorp and Majeed, 1995) have expanded this model into a two-dimensional (2-D) competition space to represent a more realistic situation in which firms compete by more than one characteristic. Thus, agents move in a 2-D characteristic space according to changes in supply (i.e. service providers) and demand (i.e. consumer preferences and needs). The model is illustrated in Figure 1, where triangles depict a platform, persons depict consumers, faces depict main service providers and stars depict complementary service providers. Each dimension represents an important characteristic of the product, and the resulting differentiation cost is defined by Euclidean distance. In other words, both characteristics are equally important, and the costs of differentiation are linear.

From a user perspective, the value of joining a platform depends on the type of data network effect, which may change from case to case. In general terms, each user will decide on switching to another platform by considering the net benefits obtained from the new platform minus the switching costs of changing from one platform to another. In the studied case, the value comes from the available data in the platform for different types of users. In general, the value of joining a platform can be expressed as follows:

$$V = U_{sa} + U_{dne} - sc - p \quad (1)$$

where V is the value of joining a platform, U_{sa} is the standalone utility or net valuation of the platform without externalities and switching costs, U_{dne} is the utility due to the data network effect, sc is the switching costs and p is the price of service subscription (per service, per platform or both). The standalone utility U_{sa} is the utility the user assigns to the platform

Figure 1 (Left) model for a horizontally differentiated area. (Right) model implementation by Repast Symphony, v 2.9



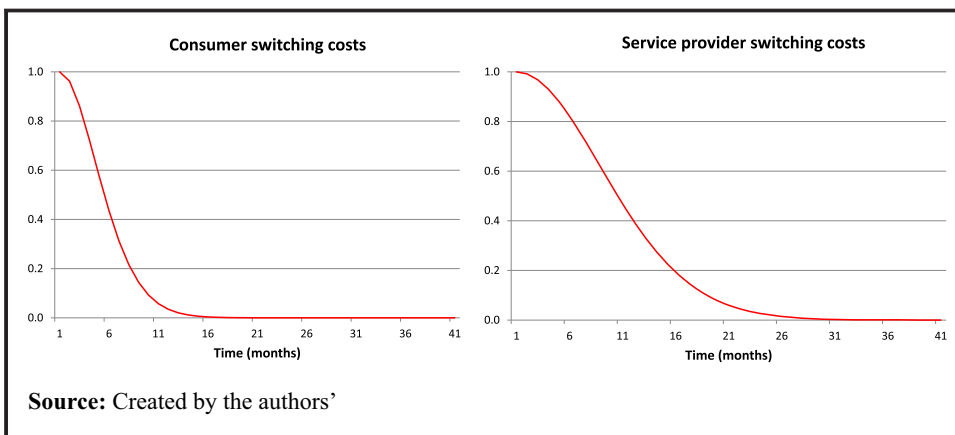
quality U_q , which is assumed to be equal for all platforms, minus the costs due to horizontal differentiation c_d (distance between the user and the platform):

$$U_{sa} = U_q - c_d. \quad (2)$$

Switching costs are different for consumers and service providers. For consumers, switching costs come from the initial investment on a connected device (device lock-in), which diminishes over time. For service providers, switching costs come from the initial investment for developing a service to be provided in a platform, which is considered a sunk cost and depreciates over time. In Figure 2, the consumer device depreciates in one year (consumer switching costs become 0), while provider switching costs assume a payback time of two years for a service (e. g. software) development by following a negative exponential function:

$$switching\ costs = e^{-\frac{k_1 t^2}{k_2}}. \quad (3)$$

Figure 2 Graphical representation of consumer and service provider switching costs (left: $k_1 = 1, k_2 = 25$; right $k_1 = 1, k_2 = 122$)



In general terms, the function describing the data network effect (i.e. data utility function) can be represented by a logistic function, in line with what is suggested by the literature (e.g. [Evans and Schmalensee, 2010](#)), such as the following:

$$\text{Data utility function} = \frac{1}{1 + \frac{a e^{-bx}}{QoS}} \quad (4)$$

In this formulation, quantity and quality are the two main attributes of data. Variable x indicates the amount of data available for a user (quantity), and QoS refers to the *quality of service* enabled by complementary data (or data quality). In addition, parameters a and b define the shape of the curve.

In particular, a consumer values the information available on services from service providers. This service information is complemented by its data history, which provides the consumer with better service offers. On the other hand, a service provider values the data on consumers so they can reach them through their services. Complementary data on other services likewise enrich their offer with better service value and consumer matches. On top of that, for a platform, the value of data consists of both consumer and service data. We can assume that the first (i.e. consumer data) relates to the quantity and the second (i.e. service data) relates to the quality of data (both data from main and complementary service providers). [Figure 3](#) shows a graphical representation of the data network effect (i.e. data utility function) for each platform user and the platform. [Table 2](#) summarizes the dimensions (quantity and quality) defining the data utility functions for each agent.

Figure 3 General representation of data utility function (i.e. value of data)



Table 2 Parameters for the utility function for different agents

Data utility function	Consumer		Service provider		Platform	
Value dimension	Quantity (x)	Quality (QoS)	Quantity (x)	Quality (QoS)	Quantity (x)	Quality (QoS)
Type of data	(All) service provider data	Own data	Consumer data	Complementary service data	Consumer data	(All) service provider data

Source: Created by the authors'

3.2 Scenarios and main parameters

The model is configured to analyze the following two scenarios:

1. Initial even market share: describes a situation where each competing platform starts with a similar amount of consumers and service providers.
2. Initial uneven market share: describes a situation where competing platforms have different initial market shares and so the initial market concentration is high. This scenario has a market with one initial dominant platform and two other challenging new entrants with considerably smaller market shares.

For each scenario, simulations run different combinations of parameters to see how they affect the performance of the simulated digital market. Each parameter represents a decision that can be promoted (directly or indirectly) by regulatory authorities, and they are described in the following [Table 3](#).

These parameters affect the data utility functions of each agent differently, as identified in [Table 4](#).

The simulations are run as explained in the following section.

Table 3 Parameters for the simulations with their values

Parameters	Values
<p><i>Device lock-in</i> Evaluates the costs the consumer faces when a device is attached to one platform. If such costs exist, they diminish with time as the device loses its value and may be replaced</p>	<p>Yes/no If yes, switching costs diminish to near zero within 12 months</p>
<p><i>Differentiation</i> Evaluates the level of horizontal differentiation between platforms. As data sharing and cooperation may decrease the level of platform differentiation, this parameter aims to test the impact of a lower differentiation</p>	<p>Normal/low Low differentiation means that the cost of distance is lower compared with a normal case (assumed to be one-third in the simulation)</p>
<p><i>Data portability</i> Describes whether data can be moved in a compatible form when a consumer or service provider changes from one platform to another</p>	<p>Yes/no If yes, data migrate when the user switches to another platform</p>
<p><i>Complementary data</i> Describes whether platforms and related service providers can use data from other complementary services to improve their offer to consumers. If the usage of complementary data is restricted, the restriction applies only to large platforms (i.e. gatekeepers)</p>	<p>Yes/no If yes, usage of complementary data is allowed. The restriction only applies to those platforms with a market share higher than 25%</p>
<p><i>Service provider multi-homing</i> Indicates the ability of service providers to reach consumers through more than one platform simultaneously</p>	<p>Yes/no When multihoming is allowed, 70% of the budget of the developed service may be used on another platform. Assuming a budget distribution of $b \sim U(a, 2a)$, with a being the required investment, 40% of the services will multihome in three platforms and 30% will multihome in two platforms. On average, a service will be provided through two platforms</p>

Source: Created by the authors'

Table 4 Parameters affecting the different dimensions of the utility functions

Data utility function Type of data	Consumer		Service provider		Platform	
	(All) service provider data (quantity)	Own data (quality)	Consumer data (quantity)	Complementary service data (quality)	Consumer data (quantity)	(All) service provider data (quality)
Device lock-in		X	X		X	
Data portability	X	X	X	X	X	X
Complementary data	X			X		X
Service provider multihoming	X			X		X

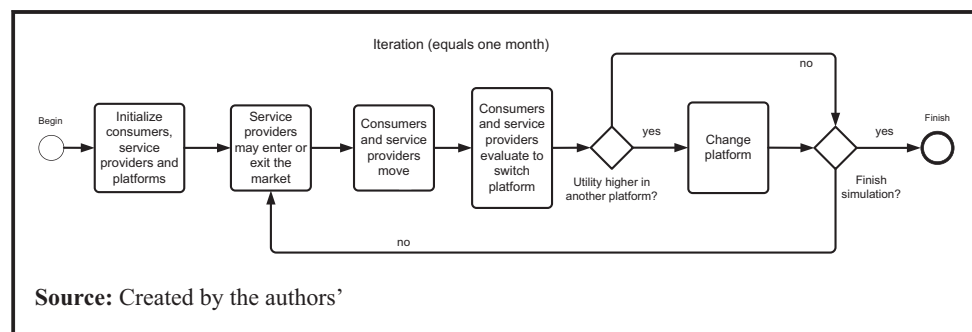
Source: Created by the authors'

3.3 Model configuration and simulation run

Each simulation represents the evolution of the market over 20 years, with each iteration describing a period of one month. For each scenario, each simulation run is repeated 100 times to obtain average values and related standard errors. The model considers 40 consumers, 20 main service providers and 20 complementary service providers. Each consumer in the simulation represents several thousand real-world consumers located in a similar niche. Since consumers' needs and tastes are dynamic, they move 10 times faster than service providers, who often lag in reacting to changes.

Figure 4 describes one simulation iteration. All agents are initialized at the beginning of a simulation run. At each iteration (which equals 1 month), service providers may exit or enter the market if the maximum amount of service providers has not been reached. After that, consumers and service providers move following a random walk. They then evaluate whether to stay on the same platform or change to another by computing the utility obtained from each platform, including the switching costs, using equations (1)–(4).

At each iteration, the total number of users is computed for each platform. At the end of one simulation (20 years), the average number of users (i.e. consumers and service providers) is calculated for each platform and the concentration index (Herfindahl-Hirschman Index, HHI) is assessed for the whole period. The HHI index value is derived by summing the squared market shares of each firm competing in a market, thus indicating the concentration of the whole market. Each simulation run is repeated 100 times for each scenario. Each repetition has the same initial parameters (including seed number), except the location of consumers and service providers, which is randomly assigned following a uniform distribution.

Figure 4 Simulation diagram describing one simulation iteration

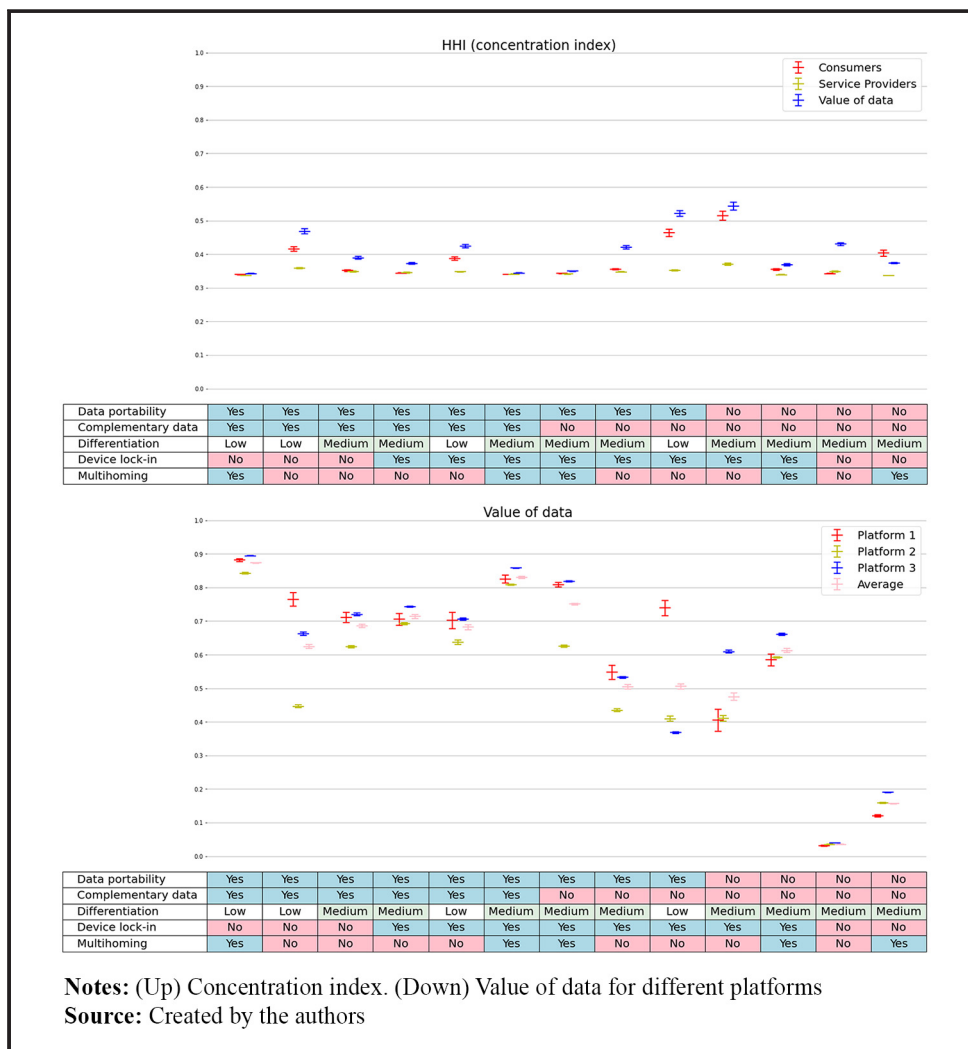
4. Results

The simulation results for the analyzed scenarios are presented in this section. For each scenario, different combinations of parameters were run. The following figures show the results for each scenario, consisting of the value of data for each platform and the concentration index (HHI) for consumers, service providers and value of data of platforms.

4.1 Initial even market share

Figure 5 shows that a combination of data portability, complementary data usage and service provider multihoming optimizes the value of data while making the market competitive. On the contrary, a combination of no data portability, restriction in the use of complementary data and no device lock-in obtain the worst results, as such a combination does not enable platforms to increase their value. The restriction of using complementary data decreases the value of data for all platforms and increases market concentration. When the lack of complementary data is combined with no service provider multihoming, both platform value and market concentration results worsen. Additionally, the lack of data portability further worsens the results. Therefore, the

Figure 5 Results for an initial even market share

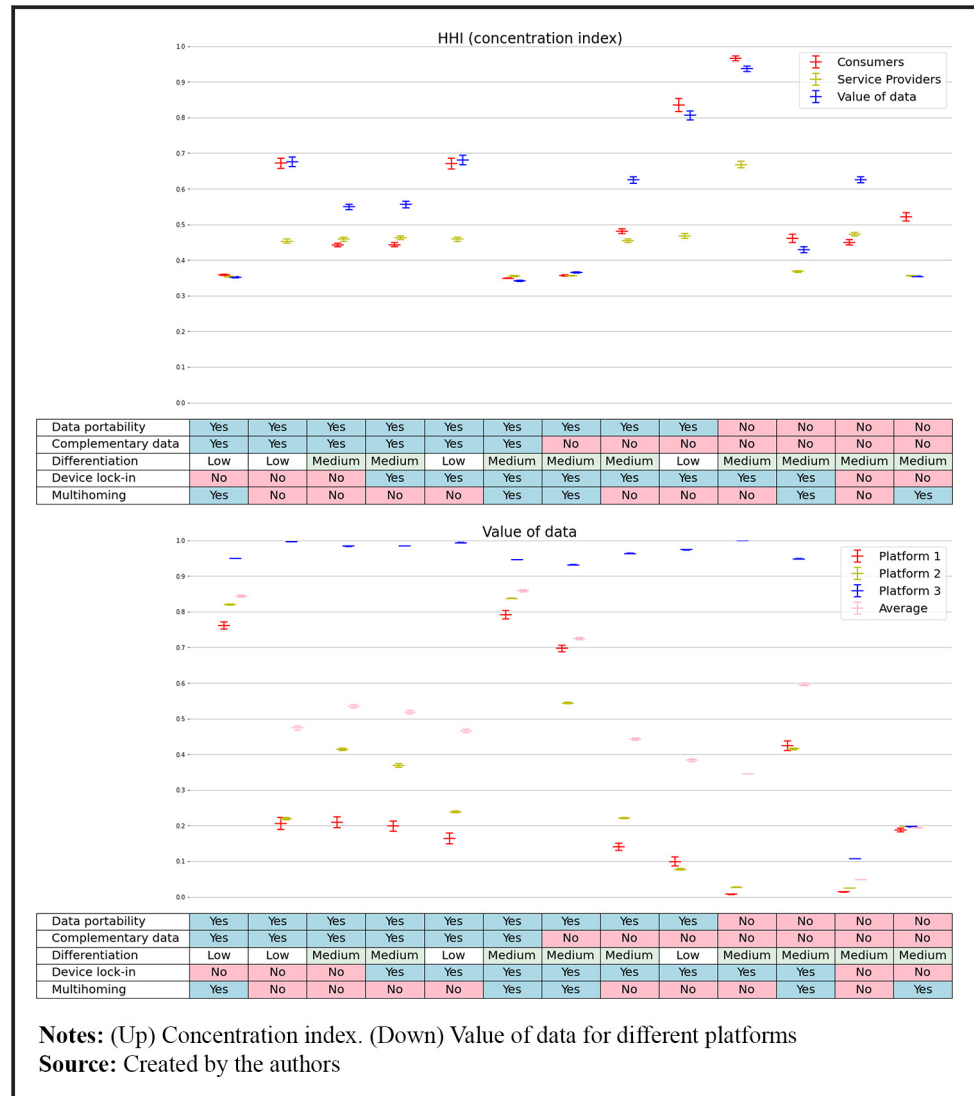


worst combination is no data portability, no use of complementary data and no service provider multihoming. No device lock-in may decrease the value in combination with no data portability and complementarity, but it has little impact on results when data portability and complementarity are allowed. Low differentiation may also slightly decrease platform and market performance, but lower differentiation may be associated with data sharing, which has a positive effect.

4.2 Initial uneven market share

Figure 6 shows how the previous combination of parameters performs for an initial high market concentration, where an initial dominant position is challenged by two smaller platforms (i.e. new entrants). In such a case, many combinations of parameters show high market concentration and low average value of data. However, optimal results are also achieved in a situation in which data portability, use of complementary data and service provider multihoming are allowed. This observation is important since this combination of parameters can correct a situation in which the market shows a dominant position. When

Figure 6 Results for an initial uneven market share



any of these three parameters is changed, platform and market performance continue to suffer high market concentration and low value of data.

As previous simulations depicted in Figures 5 and 6 show, the usage of complementary data is beneficial. In Figure 7, a sensitivity analysis of complementary data restriction with different sizes of gatekeepers (in terms of market share) is performed for a selected combination of parameters (uneven market share, device lock-in, data portability and no multihoming). The analysis shows that the restriction for using complementary data damages the market most when it is applied to all platforms. When the obligation criterion is relaxed (applied to large platforms only), the market concentration decreases and the value of the market increases. Interestingly, a high criterion for the restriction (i.e. 50% of market share) is as beneficial as not having a restriction at all. In other words, these results imply that it might be best not to apply this policy.

5. Conclusions

This study shows that a combination of data portability, use of complementary data and service provider multihoming is necessary to achieve high platform value and market competition in digital markets. It reveals that restricting data combination practices (i.e. use of complementary data) for large companies may not achieve further market contestability. Results show that even when such a restriction is applied to very large platforms, there are no clear benefits for increasing market value and competition. Other ex-ante mechanisms incentivizing data sharing and portability may effectively increase market competition and contestability. In general, this study supports the role of the regulator as a market enabler facilitating data compatibility and service multihoming.

First, data portability increases market competition and platform value. However, the usage of complementary data increases platform value and does not necessarily damage competition. On the contrary, prohibiting the use of complementary data can damage the market by favoring a dominant position given the difficulty of applying an optimal restriction criterion. Therefore, this study suggests that market contestability should be addressed through data sharing and compatibility. Complementary data may also soften network effects and make the market contestable. In general, data portability drives competition and usage of complementary data drives value. Both aspects are desirable in digital markets.

Figure 7 Concentration index and value of data results with varying levels of complementary data restriction



Additionally, service provider multihoming is a useful means to push platform competition. It is an indirect way of promoting data portability, and it works best in combination with data portability and the use of complementary data.

Very low switching costs (no device lock-in with data portability) may increase market concentration in some situations. Low switching costs (no device lock-in but without data portability) do not enable platforms to generate value for users. If platforms compete with low differentiation, then market concentration will increase, but the value of data does not suffer significantly. From this perspective, service provider multihoming may help to address the challenges related to low switching costs and low differentiation, together with data sharing and usage of complementary data.

By introducing a novel approach, this work supports previous concerns in this domain about restricting the usage of data from complementary sources. While previous studies are typically concerned with costs and other practical implementation advantages of large platforms derived from new obligations (Gal and Aviv, 2020; Baschenhof, 2022; Botta and Borges, 2023), our work emphasizes how feasible it is to achieve value from data for different market participants under different regulatory conditions. In particular, we demonstrate the effects of combining data with other regulatory mechanisms, such as data portability and service provider multihoming enabled by interoperability. Some previous works focused on the value of data (e.g. Krämer and Schnurr, 2022; Bostoen, 2023), but our study goes further by using agent-based simulations to show the dynamics of data value creation and its impact on competition and market concentration. We observe that agent-based modeling can effectively describe the dynamics of value creation based on data and can thus guide the design of data-related policies.

5.1 Practical implications and limitations

This study has practical implications for regulation. For example, it corroborates the concern regarding Article 5(2) of the DMA but also supports the obligations pushing interoperability (Articles 6(3), 6(4), and 6(5)) and data portability (Articles 5(9), 5(10), and 5(11)). According to our analysis, imposing obligations to restrict data combination practices (Article 5(2)) can damage competition. As Article 5(2) is already included in the DMA, we support that user consent should be further automatized and standardized to include all stakeholders in a fair manner, thus preventing larger platforms from obtaining any advantage.

Our work corroborates previous concerns on data combination restrictions by using a novel method to illustrate the dynamics of the effects of such restrictions from the perspective of data value. In our study, we reveal that large platforms have an advantage in obtaining value from data under data combination restrictions. This finding is in line with observations in the GDPR implementation, where compliance costs and other requirements provided an advantage to large platforms as they were more prepared to implement the required changes (Gal and Aviv, 2020). Our study provides an additional perspective to explain this advantage.

This study supports an automatized and transparent mechanism that facilitates data sharing and interoperability. Data markets can still increase transparency and user involvement. Some solutions for this have been suggested in the literature, for instance, a trust structure with visible data-sharing rules (Koskinen *et al.*, 2023), a sector-wide agreement on personalization transparency (Van Buggenhout *et al.*, 2023), trusted data intermediaries (Podszun, 2022) and bulk sharing of broad anonymized data complemented with personalized user data portability (Krämer and Schnurr, 2022). Our results, in general, indicate that regulations such as the DMA should be implemented in a way that creates clear and fair data-sharing rules (e.g. through a transparent and automatized mechanism facilitating user consent) rather than increasing restrictions that could harm value creation.

Any new obligation should be designed to be equally feasible for all stakeholders to avoid asymmetries between them. From the perspective of this work, the paradigm of data as labor (Yan and Hen, 2022), in which users actively participate in data activities in exchange for economic benefits, supports the idea of value creation by increasing transparency and clarifying market rules.

A limitation of this work is that, although agent-based modeling and simulations describe the dynamics of data markets and platform competition, they do not provide accurate forecasts on possible market outcomes. The strength of these methods is in explaining the emergent dynamics and effects of combined conditions and not in quantitatively assessing market performance. Nevertheless, the model could be further calibrated and validated once data are available after the DMA implementation.

References

- Abrahamson, Z. (2014), "Essential data", *The Yale Law Journal*, Vol. 124 No. 3, pp. 867-881.
- Bain, J.S. (1956), *Barriers to New Competition: their Character and Consequences in Manufacturing Industries*, Harvard University Press, Cambridge, Massachusetts, doi: [10.4159/harvard.9780674188037](https://doi.org/10.4159/harvard.9780674188037), ISBN: 9780674188020.
- Basaure, A., Vesselkov, A. and Töyli, J. (2020), "Internet of things (IoT) platform competition: consumer switching versus provider multihoming", *Technovation*, Vol. 90-91, p. 102101.
- Baschenhof, P. (2022), "The digital markets act (DMA): a procompetitive recalibration of data relations?", *U. III. JL Tech. & Pol'y*, Vol. 1.
- Bauer, J.M. (2022), "Toward new guardrails for the information society", *Telecommunications Policy*, Vol. 46 No. 5, p. 102350.
- Bauer, J.M. and Bohlin, E. (2022), "Regulation and innovation in 5G markets", *Telecommunications Policy*, Vol. 46 No. 4, p. 102260.
- Bostoen, F. (2023), "Understanding the digital markets act", *The Antitrust Bulletin*, Vol. 68 No. 2, pp. 263-306.
- Botta, M. and Borges, D. (2023), "User consent at the interface of the DMA and the GDPR. A privacy-setting solution to ensure compliance with ART. 5 (2) DMA", RSC Working Paper 2023/68.
- Calvano, E. and Polo, M. (2021), "Market power, competition and innovation in digital markets: a survey", *Information Economics and Policy*, Vol. 54, p. 100853.
- Cennamo, C., Kretschmer, T., Constantinides, P., Alaimo, C. and Santaló, J. (2023), "Digital platforms regulation: an innovation-centric view of the EU's digital markets act", *Journal of European Competition Law & Practice*, Vol. 14 No. 1, pp. 44-51.
- Chiou, L. and Tucker, C. (2017), "Content aggregation by platforms: the case of the news media", *Journal of Economics & Management Strategy*, Vol. 26 No. 4, pp. 782-805.
- Eaton, B.C. and Lipsey, R.G. (1975), "The principle of minimum differentiation reconsidered: some new developments in the theory of spatial competition", *The Review of Economic Studies*, Vol. 42 No. 1, pp. 27-49.
- Economides, N. (1986), "Minimal and maximal product differentiation in hotelling's duopoly", *Economics Letters*, Vol. 21 No. 1, pp. 67-71.
- Eisenmann, T., Parker, G. and Van Alstyne, M.W. (2006), "Strategies for two-sided markets", *Harvard Business Review*, Vol. 84 No. 10, p. 92.
- Evans, D.S. and Schmalensee, R. (2010), "Failure to launch: critical mass in platform businesses", *Review of Network Economics*, Vol. 9 No. 4.
- Evans, D.S. and Schmalensee, R. (2017), "Debunking the network effects bogeyman", *Regulation*, Vol. 40, p. 36.
- Finley, B. and Basaure, A. (2018), "Benefits of mobile end user network switching and multihoming", *Computer Communications*, Vol. 117, pp. 24-35.

- Gal, M.S. and Aviv, O. (2020), "The competitive effects of the GDPR", *Journal of Competition Law & Economics*, Vol. 16 No. 3, pp. 349-391.
- Geradin, D., Karanikioti, T. and Katsifis, D. (2021), "GDPR myopia: how a well-intended regulation ended up favouring large online platforms-the case of ad tech", *European Competition Journal*, Vol. 17 No. 1, pp. 47-92.
- Gregory, R.W., Henfridsson, O., Kaganer, E. and Kyriakou, S.H. (2021), "The role of artificial intelligence and data network effects for creating user value", *Academy of Management Review*, Vol. 46 No. 3, pp. 534-551.
- Hagiu, A. and Wright, J. (2023), "Data-enabled learning, network effects, and competitive advantage", *The RAND Journal of Economics*, Vol. 54 No. 4, pp. 638-667.
- Heimburg, V. and Wiesche, M. (2023), "Digital platform regulation: opportunities for information systems research", *Internet Research*, Vol. 33 No. 7, pp. 72-85.
- Hotelling, H. (1929), "Stability in competition", *The Economic Journal*, Vol. 39 No. 153, pp. 41-57.
- Huotari, P., Järvi, K., Kortelainen, S. and Huhtamäki, J. (2017), "Winner does not take all: selective attention and local bias in platform-based markets", *Technological Forecasting and Social Change*, Vol. 114, pp. 313-326.
- Jia, J., Jin, G.Z. and Wagman, L. (2021), "The short-run effects of the general data protection regulation on technology venture investment", *Marketing Science*, Vol. 40 No. 4, pp. 661-684.
- Johnson, G.A., Shriver, S.K. and Goldberg, S.G. (2023), "Privacy and market concentration: intended and unintended consequences of the GDPR", *Management Science*, Vol. 69 No. 10, pp. 5695-5721.
- Jones, C.I. and Tonetti, C. (2020), "Nonrivalry and the economics of data", *American Economic Review*, Vol. 110 No. 9, pp. 2819-2858.
- Jullien, B. and Sand-Zantman, W. (2021), "The economics of platforms: a theory guide for competition policy", *Information Economics and Policy*, Vol. 54, p. 100880.
- Kaplan, J., McCandlish, S., Henighan, T., Brown, T.B., Chess, B., Child, R., ... and Amodei, D. (2023), "Scaling laws for neural language models", *Digital Policy, Regulation and Governance*, arXiv preprint arXiv:2001.08361.
- Kennedy, J. (2017), "The myth of data monopoly: why antitrust concerns about data are overblown", *Information Technology & Innovation Foundation*, Vol. 3, pp. 19-25.
- Kobayashi, B.H. and Wright, J.D. (2020), "Antitrust and ex-ante sector regulation", *The Global Antitrust Institute Report on the Digital Economy*, Vol. 25.
- Kölbel, T. and Kunz, D. (2020), "Mechanisms of intermediary platforms", arXiv preprint arXiv:2005.02111.
- Koskinen, J., Knaapi-Junnila, S., Helin, A., Rantanen, M.M. and Hyrynsalmi, S. (2023), "Ethical governance model for the data economy ecosystems", *Digital Policy, Regulation and Governance*, Vol. 25 No. 3, pp. 221-235.
- Krämer, J. and Schnurr, D. (2022), "Big data and digital markets contestability: theory of harm and data access remedies", *Journal of Competition Law & Economics*, Vol. 18 No. 2, pp. 255-322.
- Lambrecht, A. and Tucker, C.E. (2015), "Can big data protect a firm from competition?", Available at SSRN 2705530.
- Macal, C. and North, M. (2014), "Introductory tutorial: agent-based modeling and simulation", *Proceedings of the Winter Simulation Conference 2014, IEEE*, pp. 6-20.
- Moody, D.L. and Walsh, P.A. (2002), "Measuring the value of information: an asset valuation approach", in Morgan, B. and Nolan, C. (Eds), *Guidelines for Implementing Data Resource Management*, 4th ed., DAMA International Press, Seattle.
- OECD (2021), "Ex ante regulation of digital markets", OECD competition committee discussion paper, available at: www.oecd.org/daf/competition/ex-ante-regulation-and-competition-in-digital-markets.htm
- Parker, G., Petropoulos, G. and Van Alstyne, M.W. (2020), "Digital platforms and antitrust", Bruegel, JSTOR, available at: www.jstor.org/stable/resrep50102
- Podszun, R. (2022), "Should gatekeepers be allowed to combine data? Ideas for art. 5 (a) of the draft digital markets act", *GRUR International*, Vol. 71 No. 3, pp. 197-205.

- Rubinfeld, D.L. and Gal, M.S. (2017), "Access barriers to big data", *Ariz. L. Rev.*, Vol. 59, p. 339.
- Sanchez-Cartas, J.M. (2018), "Agent-based models and industrial organization theory. A price-competition algorithm for agent-based models based on game theory", *Complex Adaptive Systems Modeling*, Vol. 6 No. 1, pp. 1-30.
- Schaefer, M. and Sapi, G. (2020), "Learning from data and network effects: The example of internet search", DIW Berlin, German Institute for Economic Research, No. 1894.
- Stigler, G.J. (1983), *The Organization of Industry*, University of Chicago Press, Chicago, IL.
- Stucke, M.E. and Grunes, A.P. (2016), "Introduction: big data and competition policy", *Big Data and Competition Policy*, Oxford University Press, Oxford, p. 1.
- Tesfatsion, L. and Judd, K.L. (Eds.). (2006), *Handbook of Computational Economics: agent-Based Computational Economics*, Elsevier, Cham.
- Tirole, J. (2023), "Competition and the industrial challenge for the digital age", *Annual Review of Economics*, Vol. 15 No. 1, pp. 573-605.
- Tucker, C. (2019), "Digital data, platforms and the usual [antitrust] suspects: network effects, switching costs, essential facility", *Review of Industrial Organization*, Vol. 54 No. 4, pp. 683-694.
- Utz, C., Degeling, M., Fahl, S., Schaub, F. and Holz, T. (2019), "(Un) informed consent: studying GDPR consent notices in the field", *Proceedings of the 2019 ACM SIGSAC conference on computer and communications security*, pp. 973-990.
- Valavi, E., Hestness, J., Ardalani, N. and Iansiti, M. (2021), "Time and the value of data", In *Academy of Management Proceedings*, Vol. 2021, No. 1, p. 13609, Briarcliff Manor, New York, NY 10510, *Academy of Management*.
- Van Buggenhout, N., Van den Broeck, W., Van Zeeland, I. and Pierson, J. (2023), "Personal data and personalisation in media: experts' perceptions of value, benefits, and risks", *Digital Policy, Regulation and Governance*, Vol. 25 No. 3, pp. 305-324.
- Veendorp, E.C.H. and Majeed, A. (1995), "Differentiation in a two-dimensional market", *Regional Science and Urban Economics*, Vol. 25 No. 1, pp. 75-83.
- Yan, X. and He, H. (2022), "Fine-tuning the ex ante approach to regulating data combination practices", *Journal of Competition Law & Economics*, Vol. 18 No. 4, pp. 881-904.

Corresponding author

Arturo Basaure can be contacted at: arturo.basaure@oulu.fi

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgroupublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com