

Krämer, Jan; Shekhar, Shiva; Hofmann, Janina

Conference Paper

Regulating Algorithmic Learning in Digital Platform Ecosystems through Data Sharing and Data Siloing: Consequences for Innovation and Welfare

31st European Conference of the International Telecommunications Society (ITS): "Reining in Digital Platforms? Challenging monopolies, promoting competition and developing regulatory regimes", Gothenburg, Sweden, 20th - 21st June 2022

Provided in Cooperation with:

International Telecommunications Society (ITS)

Suggested Citation: Krämer, Jan; Shekhar, Shiva; Hofmann, Janina (2022) : Regulating Algorithmic Learning in Digital Platform Ecosystems through Data Sharing and Data Siloing: Consequences for Innovation and Welfare, 31st European Conference of the International Telecommunications Society (ITS): "Reining in Digital Platforms? Challenging monopolies, promoting competition and developing regulatory regimes", Gothenburg, Sweden, 20th - 21st June 2022, International Telecommunications Society (ITS), Calgary

This Version is available at:

<https://hdl.handle.net/10419/265645>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Regulating Algorithmic Learning in Digital Platform Ecosystems through Data Sharing and Data Siloing: Consequences for Innovation and Welfare

Jan Krämer

University of Passau,
Dr.-Hans-Kapfinger-Str. 12,
94032 Passau, Germany
jan.kraemer@uni-passau.de

Shiva Shekhar

University of Passau,
Dr.-Hans-Kapfinger-Str. 12,
94032 Passau, Germany
shiva.shekhar@uni-passau.de

Janina Hofmann

University of Passau,
Dr.-Hans-Kapfinger-Str. 12,
94032 Passau, Germany
janina.hofmann@uni-passau.de

Abstract

Algorithmic learning gives rise to a data-driven network effects, which allow a dominant platform to reinforce its dominant market position. Data-driven network effects can also spill over to related markets and thereby allow to leverage a dominant position. This has led policymakers to propose data siloing and mandated data sharing remedies for dominant data-driven platforms in order to keep digital markets open and contestable. While data siloing seeks to prevent the spillover of data-driven network effects generated by algorithmic learning to other markets, data sharing seeks to share this externality with rival firms. Using a game-theoretic model, we investigate the impacts of both types of regulation. Our results bear important policy implications, as we demonstrate that data siloing and data sharing are potentially harmful remedies, which can reduce the innovation incentives of the regulated platform, and can lead overall lower consumer surplus and total welfare.

Keywords: Data-driven network effects, algorithmic learning, regulation, data sharing, data siloing

Introduction

Algorithmic learning gives rise to data-driven network effects (Argenton and Prüfer, 2012; Prüfer and Schottmüller, 2020; Gregory et al., 2020), which have now taken a central role in the debate on the regulation of dominant digital platforms (Kraemer and Schnurr, 2021; Parker et al., 2021; European Commission, 2020; Cennamo and Sokol, 2021). Data-driven network effects are an indirect network effect that constitutes a virtuous cycle as follows: The use of a data-driven service (or product) generates more data, which is the basis for improved algorithmic learning and data analytics, which then allows to further improve the data-driven service, and which then ultimately increases demand, which in turn generates even more data, and so on and so forth. The prime example of a service with strong data-driven network effects is a search engine (Argenton and Prüfer, 2012; Schaefer et al., 2018). As more consumers use the search

engine, more click and query data is collected allowing the search engine to improve its search algorithm, thereby drawing in more consumers, which creates even more click and query data.

In the digital economy the number and types of services that exhibit data-driven network effects, or that can be designed to exhibit data-driven network effects seems endless, and includes not only classic online services, such as search engines, social networks and media streaming (e.g., through enabling personalized recommendations). With the rise of the internet of things and artificial intelligence (e.g., face and speech recognition, image detection and text prediction), the issue of data-driven network effects is also becoming more and more prevalent and relevant in the physical sphere, as there is now a 'computer in the middle of every transaction' (Varian, 2014) that collects user data.

The virtuous cycle of data-driven network effects and algorithmic learning generates economic efficiencies and enhances consumer experience. So, why are regulators around the world concerned? First, the economic efficiencies created by data-driven network effects and algorithmic learning also give rise to economic dominance, which stifles competition and can lower incentives to innovate (Prüfer and Schottmüller, 2020; Hagi and Wright, 2020). Second, when the data gathered in one market can also be used to improve algorithmic learning and thus the quality of products and services in another market, data-driven network effects create a data externality across markets that yield a 'domino effect' whereby the related market can subsequently become dominated by the same firm (Prüfer and Schottmüller, 2020). This is because data externalities (originating from dominated related markets) make it easier to aggressively enter and win over incumbents in other related markets (Kraemer and Schnurr, 2021). For example, it is argued that Google's data and algorithmic insights on consumers' search (e.g., on locations, local businesses, opening hours) enabled it to enter the market for digital maps with a superior service (Prüfer and Schottmüller, 2020). Furthermore, as consumers installed Google maps on their mobile phone, they shared location data, which further improved its maps service with traffic information and busy hours of businesses and provided additional data that could be fed into existing and new algorithms. In turn, its superior insights from maps conveyed an advantage on Google in the autonomous driving market, which again could be used to improve algorithmic learning. Similarly, Google recently entered the health insurance market (Humer and Vengattil, 2020), based on consumer data derived from existing services, which allowed it to personalize its insurance contracts and also improve the risk models much better than standard insurance companies. In reverse, this also means that the lack of access to similar data troves, prevents algorithmic learning and data-driven network effects at the same scale. The resulting quality disadvantage can then be a major impediment for rivals to compete in platform markets (Zhu and Iansiti, 2012).

As a consequence of this, regulators have proposed two fundamentally distinct policy measures to keep digital markets competitive and to avoid long run monopolization due to data-driven network effects stemming from algorithmic learning capabilities: mandated 'data siloing' (also referred to as data separation) and mandated 'data sharing' (Bundeskartellamt, 2019; CMA, 2020; European Commission, 2020). For example, both regulatory measures could be imposed on dominant firms operating a "core platform service" (i.e., who are dominant in one data-driven market) under the newly proposed Digital Markets Act in the EU (European Commission, 2020).

Mandated *data siloing* refers to a policy intervention whereby the data-rich incumbent is prohibited from exploiting (consumer) data gathered in its primary (monopolized) market to improve algorithms and hence its services in a secondary (competitive) market. This implies that different business entities of the regulated firm (e.g. the search unit and the health insurance unit) are proscribed from sharing consumer data or insights derived thereof. In theory, this would dampen the spill over of data externalities and lower inter-market leverage of data advantage. Data siloing has already been imposed in a number of jurisdictions and cases. For example, data siloing was a remedy when approving the merger between Facebook and WhatsApp (European Commission, 2014). It is also a suggested remedy in a recent report by the British Competition and Markets Authority (CMA) on the digital advertising market (CMA, 2020). In particular, data siloing is explicitly listed as a remedy in Article 5(a) of the Digital Markets Act (DMA) (European Commission, 2020), which is the proposed legal framework for the ex-ante regulation of digital platforms in the European Union. In practice, however, data siloing will often not be perfect. First, it is easy to circumvent and difficult to monitor for regulators. Indeed, later Facebook was fined because it violated the merger remedies and combined user data from WhatsApp with other services (European Commission, 2017). Second, when data siloing is imposed as a remedy, regulators usually only demand data siloing by

default, but allow users to opt-in to the combination of their data from the various services that they are using, so that the data of those users can be exchanged nevertheless. (see, e.g. Vincent, 2019; European Commission, 2020). However, data siloing comes at a cost of reducing the efficiencies associated with the data-driven network effects between markets (Kraemer and Schnurr, 2021).

By contrast, mandated *data sharing* allows the regulated firm to use its data acquired in the primary market also in the secondary market, but requires to share this data with rivals so that they compete on more equal footing. Mandated data sharing has been suggested as a promising policy intervention by several scholars (e.g., Argenton and Prüfer, 2012; Kraemer and Schnurr, 2021; Prüfer and Schottmüller, 2020; Parker et al., 2021), as it immediately exploits the non-rivalry of data, which is the main economic property that makes data inputs distinct from physical inputs. Thus, in contrast to data siloing, which limits the extent of data externalities and algorithmic learning, as well as its associated efficiencies, data sharing permits these efficiency spillovers to the secondary market at the cost of increased competition arising from sharing these efficiencies with the rival. Accordingly, mandated data sharing has also been suggested by the CMA (2020), and for the specific context of click and query data from search engines, it is also included under Article 6(j) of the DMA (European Commission, 2020). However, such a remedy has technical constraints. Data sharing can never be achieved fully in practice due to technical and legal limitations. Consumer data is collected at a fine granular and individual level. To establish a level playing field, this data would need to be shared instantaneously when being collected, and at the same level of detail. This is practically infeasible in most applications, however, for example due to the sheer amount of data (and the possibly many recipients of it) that accrues¹ and due to privacy regulation. Regarding the latter, both technological measures (such as federated learning and differential privacy) as well as institutional measures (such as third-party data trusts, or in-situ rights to run own algorithms on the regulated platforms data infrastructure) can be taken to overcome some of the privacy concerns (for a more detailed discussion see Kraemer and Schnurr, 2021; Parker et al., 2021), but all of these approaches yet involve limitations to the detail of the data that can be shared. Moreover, we formally show an additional, more nuanced negative consequence of such a remedy. While mandated data sharing does increase the level of competition in secondary market, it lowers the incumbent's incentive to innovate in the primary market.

Despite the practical relevance of data siloing and data sharing for the regulation of data-centric markets and online platforms, the impact of two approaches on innovation incentives have not yet been contrasted in a formal framework. This paper employs a game theoretic model to study and compare the effect of data sharing as well as data siloing on competition, innovation and welfare. Thereby, we contribute to the emerging literature analyzing the economic effects of data-driven network effects (e.g. Gregory et al., 2020; Haftor et al., 2021; Argenton and Prüfer, 2012; Schaefer et al., 2018), and the associated regulation of digital platforms (see, e.g., Hagiu and Wright, 2020; Prüfer and Schottmüller, 2020; Parker et al., 2021; Tucker, 2019; Kraemer and Schnurr, 2021). We build on prior research that has shown that data-driven network effects can be a powerful source of market power and dominance (Schaefer et al., 2018; Prüfer and Schottmüller, 2020), and that access to large data sets and improved predictions can boost innovation (e.g., Agrawal et al., 2018; Bajari et al., 2019). In particular, we adopt the notion that user experience in digital platform ecosystems strongly depends on the scale of learning that can be achieved from acquired consumer data (Gregory et al., 2020; Hagiu and Wright, 2020). In this context, not only the amount of data, but also its structure and context matter, and together give rise to the strength of the data-driven externality (Afuah, 2013; Shankar and Bayus, 2003).

In our model, we therefore consider the size of the data and the strength of the externality to be mutually reinforcing when determining the be the effect of varying degrees of data siloing and data sharing, as opposed to technically and often legally infeasible perfect data sharing, which has been considered in previous literature (Prüfer and Schottmüller, 2020; Hagiu and Wright, 2020). Our results suggest that data siloing can potentially be a harmful remedy, because it always reduces the innovation in the monopolized primary market and yields a lower (consumer) surplus in our model compared to a *laissez-faire* no regulation approach. Our model also shows that data sharing yields lower innovation in the primary market while increasing competition in the secondary market. Surprisingly, when data externalities are large consumer surplus is lower under sharing than no regulation. Nevertheless, a data sharing remedy always yields more innovation and surplus than a strict data siloing remedy.

¹For example, it is estimated that Google receives approximately 70.000 search requests per second.

Our model addresses an overlooked issue of data externalities from a primary platform market to a secondary platform market: Data externalities not only benefit the consumers in the secondary market, but also encourage innovation in the primary market. Increased innovation in the primary market expands demand in the primary market, which then fuels the virtuous cycle of data-driven network effects, and the associated data-driven advantage in the secondary market. Data siloing switches off this data externality and thereby lowers innovation incentives of the dominant platform—not only in the primary market, but also in the secondary market. This can ultimately hurt consumers in both the primary and the secondary market. By contrast, data sharing preserves the data externality, but takes away the data incumbent’s data-driven advantage in the secondary market, which lowers the incumbent’s incentive to collect data in the primary market. This means that also under a data sharing regulation the incumbent has less incentives to innovate in the primary market which spills over to lower innovation incentives of the incumbent in the secondary market as well.

While more data sharing always benefits consumers in the secondary market (as this market becomes more competitive and data efficiencies are shared), more data siloing always hurts consumers in the secondary market (as data efficiencies are denied). Overall, a combination of data siloing and data sharing regulation can therefore lower consumer surplus either in only one market (the primary market) or both markets, depending on the imposed level of data siloing and data sharing. Even if the welfare effect in the secondary market is positive, we find that the unambiguously negative welfare effect in the primary market can be stronger than the positive welfare effect in the secondary market. In summary, we show under which conditions data sharing or data siloing, or a combination thereof, can yield lower consumer and total welfare than in the case where no such regulation had been imposed.

The Model

We consider a game-theoretic model with two platforms $i = 1, 2$ and two markets $m = A, B$. In the primary market A platform 1 is a monopolist, whereas in the secondary market B platform 1 competes with another platform 2. Let q_{mi} be the demand (number of consumers) of platform i in market m . Thus, the total demand in market A is $Q_A = q_{A1}$ and in market B it is $Q_B = q_{B1} + q_{B2}$. We assume a Cournot-type competition in each market, and the inverse demand function faced by firm i in market m is

$$a_{mi} = 1 + V_{mi} - 3Q_m.$$

Thereby, V_{mi} is the quality of each platform’s service in a given market, and a_{mi} is the (implicit) price of the service. For example, one can think of a_{mi} as the consumers’ average willingness to pay for the service, or the level of advertisements that consumers would be willing to accept when using the service. Note that this inverse demand function is consistent with a number of microfoundations as in Katz and Shapiro (1985).²

We allow platforms to innovate and thus, to increase the quality of their service in each market. Specifically, platform 1 can innovate in its primary market A and in the secondary market B and thereby enhances the service quality of its products by v_{A1} and v_{B1} respectively, while platform 2 can innovate in market B and enhances its service quality by v_{B2} .

²For example, assume that consumers’ utility is given by $U_{mi}(r) = r + V_{mi} - a_{mi}$, where r is the willingness to pay for the service provided by a given platform, net of the platform’s service quality and price. Consumers will visit only the platform that provides them with the highest utility. In equilibrium, it must therefore be that the quality-adjusted price implied by each platform $\Phi_m = a_{mi} - V_{mi}$ is identical for all platforms in a given market m , as otherwise one platform would not be visited and would have an incentive to lower its implicit price, a_{mi} , in order to receive positive demand. Furthermore, we assume the value of a consumer’s outside option is zero, such that consumers with $U_{mi} < 0$ will not choose any platform. This implies total demand is constituted only by those consumers for whom $r > \Phi_m$. Assume further that there is a unit mass of consumers which differ in r , and r is uniformly distributed with support on $[-2, 1]$. As in Katz and Shapiro (1985), the support also includes negative values, as some consumers may not use the service, even at a zero price, unless the service quality exceeds a certain threshold. This support assures that there is no corner solution and that the axioms of probability are satisfied when firms invest in value enhancement of their product in either market. Then, total demand in market m is given by $Q_m = 1 - F(\Phi_m) = 1 - (a_{mi} - V_{mi} + 2)/3$. Solving for a_{mi} yields precisely the inverse demand function stated above.

In order to capture the positive data externalities from market A to market B , we assume that platform 1 can leverage its consumer data obtained in market A to further increase its service quality in market B (e.g., because the data allows it to identify relevant design features and consumer needs). Thus, the aggregate service quality in market B depends also on the demand in market A , i.e., we assume that more (or more representative) consumer data can be obtained from a larger user base in market A . In addition, the service quality in market B depends on the strength of the data externality, e.g., how useful the data obtained in market A is for improving the quality in market B . Let $\theta > 0$ denote the strength of the data externality, i.e., the marginal quality improvement in market B that can be attributed to each consumer (and its associated data) in market A .³ Taken together, the aggregate quality V_{mi} , for each service is given by

$$V_{A1} = v_{A1}, \quad V_{B1} = v_{B1} + (1 - \rho)\theta q_{A1}^e, \quad V_{B2} = v_{B2} + (1 - \rho)\theta\delta q_{A1}^e. \quad (1)$$

Thereby, we denote the degree of data siloing by $\rho \in [0,1]$, where $\rho = 0$ represents no obligation for data siloing, and $\rho = 1$ corresponds to a regime with strict data siloing obligations. Likewise, $\delta \in [0,1]$ denotes the share of the data that platform 1 is required to share with platform 2. By varying ρ and δ , we can analyze the effect of various degrees of data siloing and data sharing, also in combination with each other. For the parameter constellation $\rho = \delta = 0$, the platforms are not regulated, i.e., platform 1 can make use of all of its data from market A in market B (no data siloing) and does not have to share data with the rival platform (no data sharing).

Note that we do not consider data-driven network effects within market A or market B , but purely focus on data externality spillovers between markets and the impacts of the regulation thereof. Incorporating additional data-driven network effects within each market would qualitatively yield the same results at the cost of rendering our model less parsimonious and its insights less comprehensible. Furthermore, note that we assume that consumers make their consumption decisions in markets A and B at the same time (see timing below). When doing so, due to the presence of data externalities from market A to market B , consumers have to form expectations about the demand in market A when considering with which platform to affiliate in market B . We therefore denote the expected demand in market A by q_{A1}^e . In equilibrium, these expectations need to be fulfilled, such that $q_{A1} = q_{A1}^e$ and we will solve the game accordingly.

Finally, let $C(v)$ be a platform's service-specific cost of innovation that enhances quality of their product by a value commensurate to v . We assume a standard convex cost function, which means that it becomes increasingly costly for any platform to innovate in order to increase its service quality. For simplicity, let $C(v) = v^2/2$, but our insights do not depend on this specific functional form. Consequently, platforms' profits are given by

$$\Pi_1 = a_{A1}q_{A1} + a_{B1}q_{B1} - C(v_{A1}) - C(v_{B1}), \quad \Pi_2 = a_{B2}q_{B2} - C(v_{B2}). \quad (2)$$

The timing of the two-stage game is as follows: In the first stage, platforms choose their level of innovation v_{A1} in their respective markets. In the second stage, observing innovation levels, consumers in market B form beliefs on the expected demand, q_{A1}^e , in market A . Platforms simultaneously compete to serve consumers in all markets. Then demand and profits are realized. We solve the game employing the Fulfilled Expectations Cournot Equilibrium (FECE) concept.

The Impact of Data Siloing and Data Sharing on Market Outcomes and Welfare

Equilibrium Analysis

We solve the game backwards and consider the effect of data siloing (i.e., a change in ρ) and data sharing (i.e., a change in δ) on equilibrium outcomes.

³ To ensure that the model is well-behaved and second order conditions are satisfied, we assume that the data-driven network effects are not too dominant relative to the costs of direct quality investments and thus, the rival platform is active in the market also without data sharing. This is ensured by assuming $0 < \theta < 1$.

In Stage 2, platforms determine simultaneously which level of demand they seek to achieve in each of the two markets considering the belief of consumers regarding the demand in market A. In order to achieve the equilibrium demands, we first solve the first order conditions simultaneously, which yields:

$$q_{A1} = \frac{1+v_{A1}}{6}, \quad q_{B1} = \frac{(1+2v_{B1}-v_{B2}+\theta(2-\delta)(1-\rho)q_{A1}^e)}{9}, \quad q_{B2} = \frac{(1+2v_{B2}-v_{B1}-\theta(1-2\delta)(1-\rho)q_{A1}^e)}{9}. \quad (3)$$

The demand served in market A solely depends on the service quality v_{A1} and, intuitively, increases with it. In market B where consumers benefit from the data externalities, the demands served depend on consumer expectations of the final demand in market A (i.e. the size of the data), as well as the strength of the data externality, θ , and the degree of the regulatory interventions. As consumers' expectations get more favorable about the demand served in market A, it is also intuitive that the demand served by platform 1 in market B rises. Instead, the demand served by platform 2 rises with q_{A1}^e , if and only if data sharing is sufficiently large ($\delta > 1/2$). As the degree of data sharing increases, platform 2 is able to compete on a more equal footing and thus able to obtain a larger demand. In equilibrium, consumers expectations need to be fulfilled. Thus, we can set $q_{A1} = q_{A1}^e$ from now on, which yields

$$q_{B1} = \frac{(6+12v_{B1}-6v_{B2}+\theta(2-\delta)(1-\rho)(1+v_{A1}))}{54}, \quad q_{B2} = \frac{(6+12v_{B2}-6v_{B1}-\theta(1-2\delta)(1-\rho)(1+v_{A1}))}{54}. \quad (4)$$

The above expressions are intuitive. Platform demands in market with B rise with own investments and fall with the investment levels of their rival. Due to the spillover of data externality from market A to market B, demands in market B also are impacted by investments in market A. With a higher service quality (v_{A1}) in market A (which results in a larger demand in market A), or a larger importance of the data externality (θ), platform 1 obtains a larger demand in market B. Instead, platform 2 benefits from a higher v_{A1} or θ only under a sufficiently high degree of mandated data sharing i.e., $\delta > 1/2$. In reverse, when the level of data sharing is low (or completely absent) a large θ or v_{A1} reduces the demand for platform 2, because it cannot effectively compete with platform 1 due to its data disadvantage. Moreover, notice that as the level of data sharing or data siloing increases, platform 1 finds it profitable to reduce the demand it serves. While platform 2 always increases demand with data sharing, an increase in data siloing makes it profitable for platform 2 to increase demand only when data sharing levels are low i.e., $\delta < 1/2$. Interestingly, the rise in consumer demand served by platform 2 from increased data sharing is higher than the fall in consumer demand served by platform 1. This suggests that consumers benefit from data sharing in market B, given a fixed level of innovation in market A. However, a key feature of our model is that the level of innovation in market A is not fixed and chosen endogenously by the regulated firm.

In Stage 1, each platform sets innovation levels to maximizing profits. To understand better the following results, it is worthwhile to discuss the best response functions of the platforms' innovation levels in market B, which are given by:

$$v_{B1}(v_{B2}, v_{A1}) = \frac{2(6-6v_{B2}+(1+v_{A1})\theta(1-\rho)(2-\delta))}{57}, \quad v_{B2}(v_{B1}, v_{A1}) = \frac{2(6-6v_{B1}-(1+v_{A1})\theta(1-\rho)(1-2\delta))}{57}.$$

From the above, it is straightforward to see that innovation levels in market B are strategic substitutes. This means that, a unilateral *increase* in the level of innovation in market B by one platform is responded by a *reduction* in innovation by the rival platform. However, due to the data externality, an increase in innovation in market A encourages platform 1 to innovate more in market B as well. In that sense, platform 1's innovation in market A is a complement to the innovation of platform 1 in market B. Instead for platform 2, an increase in innovation by platform 1 in market A increases its innovation incentive only when $\delta > 1/2$. These best responses of the platforms aid us in explaining the nuanced results in the following proposition.

Maximizing each platforms' profit with respect to quality levels immediately yields:

Proposition 1 [Innovation] The levels of innovation in market A and market B are given as

$$v_{A1} = \frac{345+(2-\delta)(1-\rho)(30+\theta(14-9\delta)(1-\rho))}{1725-(2-\delta)(14-9\delta)(1-\rho)^2\theta^2}, \quad v_{B1} = \frac{12(25+\theta(14-9\delta)(1-\rho))}{1725-(2-\delta)(14-9\delta)(1-\rho)^2\theta^2},$$

$$v_{B2} = v_{B1} - \frac{4(1-\delta)(1-\rho)\theta(69+\theta(1-\rho)(2-\delta))}{1725-(2-\delta)(14-9\delta)(1-\rho)^2\theta^2}.$$

The level of innovation by platform 1 in market A and B is reduced with stricter data sharing (δ increases) or with stricter data siloing (ρ increases). The innovation level of platform 2 always rises with data sharing, but rises with data siloing only when the level of data sharing is low ($\delta < 9/14$).

With stricter data siloing or stricter data sharing obligations platform 1 finds it less profitable to serve consumers in market A and reduces its service quality, v_{A1} . This is because its investments in innovation in market A do not bear the same competitive advantage in market B than without the regulatory interventions. In turn, the lower demand in market A also reduces the data available for quality improvement in market B. Likewise, innovation by platform 1 in market B also falls with data sharing and data siloing. This result is a combination of two reinforcing effects. First, an increase in data sharing or siloing makes each innovation unit in market A less valuable in market B due to competition, which in turn lowers the incentive to innovate in market B. Second, an increase in data sharing or data siloing also reduces the absolute investment levels in market A which further lowers incentive to innovate in market B.⁴ This reduced innovation incentive of platform 1 affects platform 2's innovation levels in a more nuanced way. Platform 2's increased innovation incentives as data sharing increases is the sum of direct and indirect effects. The direct effect of an increase in data sharing ceteris paribus enhances innovation incentives of platform 2 as its marginal revenues rise with more data sharing. The indirect effect arises from reduced incentives to innovate by platform 1 in market A and B. While a reduction in innovation by platform 1 in market B positively impacts innovation incentive of platform 2, a fall in innovation in market A may lower incentive to innovate only when data sharing levels are sufficiently high. In sum, the positive effect dominates and platform 2 innovates more with data sharing.

Interestingly, an increase in data siloing increases the innovation level of platform 2 in market B only when the imposed level of data sharing is relatively low ($\delta < 9/14$). If the level of data sharing is high, then data siloing is detrimental to platform 2's innovation levels as the negative impact arising from lower innovation levels in market A dominates any positive innovation effect stemming from reduced investment by platform 1 in market B.

Next, we consider how data sharing and data siloing affect the competition in the two markets. To this end, we consider how the advertising intensity (i.e., the implicit price paid by consumers) is affected.

Proposition 2 [Advertising intensity] The equilibrium advertising intensities are given by

$$a_{A1} = \frac{15(69+(2-\delta)(1-\rho)\theta)}{1725-(2-\delta)(14-9\delta)(1-\rho)^2\theta^2}, a_{B1} = \frac{27(25+\theta(14-9\delta)(1-\rho))}{1725-(2-\delta)(14-9\delta)(1-\rho)^2\theta^2},$$

$$a_{B2} = \frac{9(1-\delta)(1-\rho)\theta(69+(2-\delta)(1-\rho)\theta)}{1725-(2-\delta)(14-9\delta)(1-\rho)^2\theta^2} + a_{B1}.$$

The advertising intensity of the regulated platform 1 in market A and B falls with more data sharing (δ) or more data siloing (ρ). The advertising intensity of the rival platform 2 rises with δ and falls with ρ if $\delta > 9/14$ (and rises otherwise).

Platform 1's advertising intensity in markets A and B decreases with the level of data sharing, δ , albeit for different reasons. In market A, advertising intensity falls because the service quality in market A is reduced as more data has to be shared. The lower service quality means that consumers are less willing to accept advertisements. In market B, an increase in data sharing makes the rival a fiercer competitor. This forces platform 1 to lower its advertising intensity. In contrast, the advertising intensity of platform 2 increases with data sharing, as it can now offer a more attractive service to consumers attributed to increased investment by platform 2 and increased data externality.

An increase in ρ (more restrictive data siloing) decreases platform 1's advertising intensity in both market A and B, but again for different reasons. More data siloing yields the regulated firm to innovate less in service quality in market A, which then lowers its investments in market B as well and thus the advertising level in market A falls.⁵ In market B, more data siloing leads to two reinforcing effects. First, a direct effect of a decrease in value of consumers from data-driven benefits as well as innovation effects in market B, and

⁴ See the discussion on best responses to innovation levels before Proposition 1 above.

⁵ Recall from our previous discussion of best response functions that innovation incentives of platform 1 in market B are complemented by innovation levels in market A.

second, an indirect effect arising from decreased consumer demand in market A due to lower innovation in market A . In contrast for platform 2, when data sharing is low ($\delta < \frac{9}{14}$), the data disadvantage of platform 2 is large. Hence, a more lenient data siloing regulation (a decrease in ρ), only exacerbates the data disadvantage, requiring platform 2 to lower its investments and thus also its advertising intensity.

Welfare

We measure welfare in three different ways: (1) The platforms' profits (producer surplus), (2) consumer surplus, and (3) total welfare, i.e., the sum of producer surplus and consumer surplus. By substituting the equilibrium values from Propositions 1 and 2 in the profit expression from equation (2), and performing comparative statics, we derive the following proposition.

Proposition 3 [Platforms' profits] Platform 1's profit falls with stricter data sharing (δ increases) or stricter data siloing (ρ increases). Platform 2's profit rises with δ and falls with ρ if $\delta > \frac{9}{14}$ (and rises otherwise).

Intuitively, platform 1 makes the highest profit in the absence of any regulation, that is when there are no restrictions on data siloing ($\rho = 0$) and no obligations to share data with the rival ($\delta = 0$). In reverse, the rival platform benefits from data sharing. Interestingly, the rival platform is also hurt from stricter data siloing obligations if at the same time a high degree of data sharing is imposed ($\delta > \frac{9}{14}$). By contrast, in the absence of data sharing obligations (or if the degree of data sharing is low), then a stricter data siloing obligation can indeed increase the profit of the rival platform. This highlights that, if the goal of data-specific regulations are to make rival firms better off, then policymakers need to consider the interaction of data siloing and data sharing more carefully, instead of viewing them as independent of each other.

The consumer surplus CS in each market is given by $\frac{Q_j^2}{2}$, with $Q_j = \sum_i q_{ji}$. Inserting the equilibrium values in the equations for q_{A1} in (3) and q_{Bi} in (4) and performing comparative statics, we obtain the following proposition.

Proposition 4 [Consumer surplus] A stricter data siloing regulation (an increase in ρ) decreases consumer surplus in market A and B , respectively. A stricter data sharing regulation (an increase in δ), lowers consumer surplus in market A and increases consumer surplus in market B . The total consumer surplus increases with data sharing (and increase in δ).

The intuition with respect to a change in ρ is straightforward. A stricter data siloing regime decreases innovation in market A which immediately lowers consumer surplus in market A . In market B an increase in ρ results in two reinforcing and opposing effects. First, a direct consumer welfare loss from a decrease in the data externality as well as the welfare loss from a decrease in innovation in market A , which yields less demand in market A , and eventually a smaller data externality for consumers in market B . This negative effect is further reinforced by a reduction in innovation by platform 1 in market B . Finally, platform 2 increases its innovation with data siloing only when data sharing levels are low else the negative effect on consumer surplus is further reinforced. The total effect is unambiguously negative on consumer surplus. Thus, a stricter data siloing regime stands to hurt consumers in both markets.

The impact of an increase in data sharing on consumer surplus is also a composite of multiple effects. An increase in data sharing lowers innovation in market A which reduces consumer surplus in market A . In market B , more data sharing has two opposing effects on consumers surplus. First, a direct positive effect arising from increased investments by platform 2 in market B . This makes platform B a fiercer competitor along with higher value for consumers. Second, data sharing negatively impacts consumers surplus through lower investments platform 1 in market A and market B . The positive effect of increased competition and higher investments by platform 2 in market B dominates and thus, data sharing increases consumer surplus in market B . Interestingly, the positive effect in market B from stricter data sharing can dominate the negative impact on market A (if at the same time, the level of data siloing is relatively low). This suggests that the consumer surplus loss in market A from reduced innovation may be outweighed by the consumer

surplus gains in market B , rendering a data sharing regulation potentially beneficial to consumers overall.⁶ As data siloing and data sharing have opposing effects on consumer welfare, we discuss the policy implications of a combined data sharing and data siloing regulation further below.

When considering total welfare, i.e, the sum of the platforms' profits and total consumer surplus, we obtain the following proposition.

Proposition 5 [Total Welfare] A stricter data siloing regulation (increase in ρ) decreases total welfare. A stricter data sharing regulation (increase in δ) increase total welfare when $\rho > \max\{0, \rho^W(\delta)\}$.⁷

Unsurprisingly, the effect of data siloing and data sharing reflects a combination of the aforementioned effects on producers and consumer surplus. It highlights once again, that a stricter data siloing regime, which limits the extent of the data externality can be harmful as it reduces welfare. Likewise, also a stricter data sharing regime can reduce total welfare, albeit it unambiguously increases consumer surplus in market B , as well as platform 2's profit in market B .

Conclusions and Policy Implications

Digital platform ecosystems typically exhibit strong data-driven network effects which originate from their algorithmic learning capabilities. These allow the platform not only to maintain a dominant position in its primary market, but also to leverage its superior access to consumer data in order to gain a competitive advantage in a secondary market. Although such data-driven externalities stemming from algorithmic learning bear efficiencies and allow to improve the quality of the knowledge-intensive services offered, policymakers around the world are concerned with the economic power and distortion of competition that comes about with data-driven externalities and the superior ability to derive algorithmic insights. Therefore, several policymakers in Europe have suggested to impose both data siloing obligations (which seek to limit the extent of data externalities across markets) as well as data sharing obligations (which seek to share the data externalities with rivals) for dominant platforms (European Commission, 2020; CMA, 2020).

We have developed a game theoretic model to analyze the effect of these obligations on market outcomes, competition and welfare, including the so far neglected perspective on how these interventions would affect innovation incentives of the regulated platform in the primary market. In practice, the degree to which data siloing and data sharing can be implemented is constrained by technical, legal and political limitations. For example, to which extent data can be shared with rivals depends, among other things, on the privacy rights and intellectual property rights in the relevant jurisdiction, the amount of data to be shared, the effectiveness and efficiency of privacy-preserving technologies, as well as political willpower.

In order to summarize our results and main insights, it is useful to consider and compare three stylized regulatory regimes, where policymakers either impose (i) strict data siloing (i.e., $\rho = 1$), so that neither the regulated platform, nor the rival platform can use data from the primary market for their operations in the secondary market⁸; (ii) data sharing (without strict data siloing)⁹, where $\rho \in [0,1)$ and $\delta \in (0,1]$; and (iii) no regulation, so that neither data siloing nor data sharing is imposed (i.e, $\delta = 0$ and $\rho = 0$). From Propositions 1 to 5, we can then immediately establish the following insights.

Insight 1 [Impact on innovation in the primary market] The level of innovation in the regulated primary market is higher under a data sharing regime than under the strict data siloing regime, but always highest without regulation.

Insight 2 [Impact on innovation in the secondary market] For the regulated platform 1, the level of innovation in the secondary market is higher under a data sharing regime than under the strict data siloing regime. However, innovation by platform 1 is always highest without regulation.

⁶ We note that this result also arises from our assumption that consumers are drawn from the same distribution and market size is unity. The above results are expected to be more nuanced once we allow for market size to differ.

⁷ For brevity, we do not present the expression for $\rho^W(\delta)$.

⁸ Note that it follows immediately from Propositions 4 and 5 that an intermediate degree of data siloing can never be optimal for consumer surplus or total welfare.

⁹ Note that data sharing cannot be imposed in combination with strict data siloing (i.e., $\rho = 1$).

For the rival platform 2, the level of innovation is higher under a data sharing regime than under the strict data siloing regime only if the level of data sharing regulation is high ($\delta > 9/14$). However, innovation by platform 2 is always higher with regulation than without regulation.

Insight 3 [Impact on profits] The profit of platform 1 is higher under the data sharing regime than under the strict data siloing regime, but always highest without regulation. The profit of platform 2 is higher under the data sharing regime than under the strict data siloing regime if $\delta > 9/14$, but always the lowest without regulation.

Insight 4 [Impact on welfare] In comparison to no regulation, data sharing regulation increases consumer welfare or total welfare only if the level of data siloing is low ($\rho < \rho^{CS}$ or $\rho < \rho^W$). [See Figure 1.] A strict data siloing regime always yields a lower (consumer) welfare than no regulation.

Taken together, our results bear several important policy implications. First, our analysis shows that a strict data siloing obligation can potentially be a very harmful remedy that reduces (consumer) welfare and leads to the least innovation in the primary market - even below the level without regulation. Data siloing destroys the efficiencies associated with data externalities and cuts the positive feedback loop between the primary market and the secondary market that incentivizes the dominant platform to pursue a high level of innovation. Data siloing is effective insofar, as it increases the competitiveness of the rival platform in the secondary market and leaves it with higher profits than without regulation. Arguably, data siloing is also easier to implement and to administer than a data sharing obligation, which makes it a seemingly attractive remedy to policymakers. However, our results suggest that the ease of implementation should not be a guiding principle for policymakers, and that a hands-off approach may even be better after all. In the same spirit, our results also suggest that – if policymakers want to impose data siloing obligations nevertheless due to political pressure – a more lenient data siloing provision (e.g. an opt-out rather than the currently used opt-in regime) is to be preferred.

Second, our results confirm that a data sharing obligation can increase (consumer) welfare compared to no regulation. However, this also comes with several trade-offs and caveats. Although a higher degree of data sharing is unambiguously positive for (consumer) welfare in the secondary market, because it preserves the value generated by data externality and shares this efficiency with the rival, it is detrimental to innovation and (consumer) welfare in the primary market. The overall assessment depends on the strength of the data externality that is permitted by data siloing and the amount of data sharing mandated. Data sharing obligations can only be beneficial if there are strict data sharing obligations in tandem with a lenient data siloing regime (see Figure 1).

Third, as an immediate consequence of this, policymakers need to be aware that data siloing and data sharing remedies will interact and thus, should not be considered in isolation: The higher the level of data siloing, the less data can be shared. It seems that currently policymakers are willing to impose much stricter data siloing provisions than data sharing provisions. For example, the Digital Markets Act (European Commission, 2020) foresees data sharing only for regulated search engines, whereas data siloing is imposed on all regulated platforms. In fact, our results suggest that the opposite would be more favorable for increasing welfare, i.e., imposing a less stringent data siloing regime in tandem with more comprehensive data sharing obligations.

Finally, we point to some limitations and avenues for future research. We build a parsimonious model in order to highlight the effect of mandated data siloing and data sharing between data-driven externalities and innovation between two related markets. Like any theoretical model, our model can be criticized based on its assumptions and the focus of the analysis, which led us to ignore additional effects. Indeed, our model could be extended in a number of ways. First, as mentioned above, we could incorporate data-driven network effects not only across markets, but also within each market. This would amplify the benefits of attaining a larger market share and in consequence lead to a higher level of innovation in market A as well as a stronger degree of competition in market B. However, this would not change our insights qualitatively. With mandated data siloing or data sharing, the incentives to innovate in market A would still be lower, and the degree of competition in market B would still be higher than without these regulatory interventions.

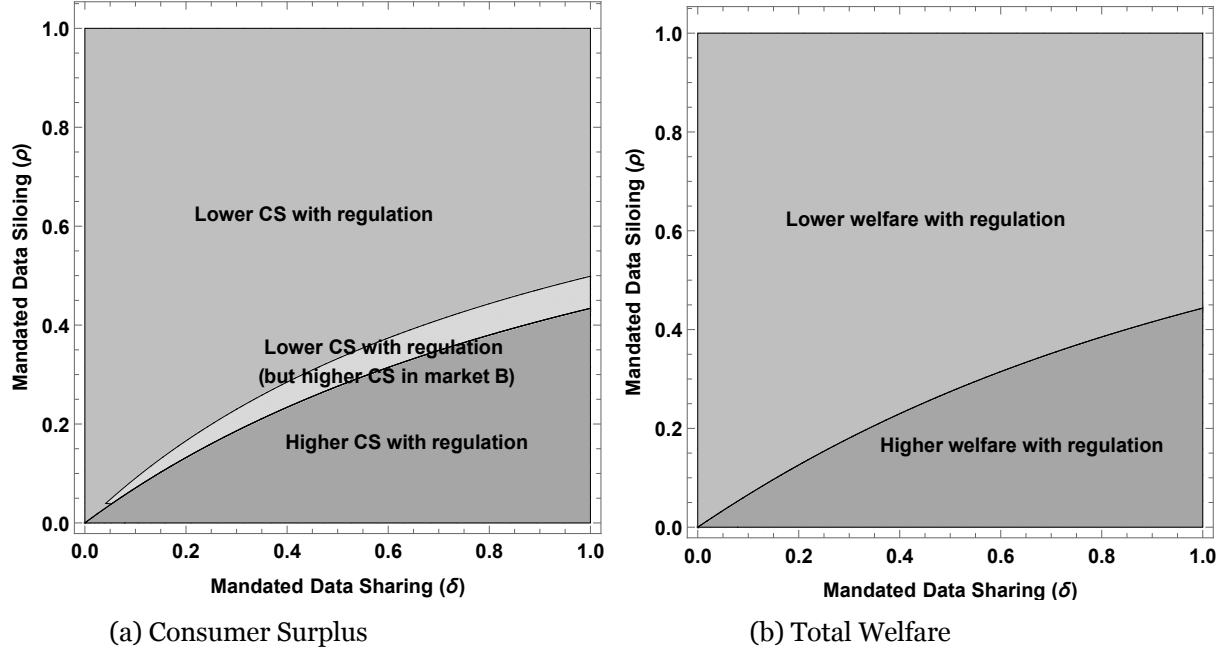


Figure 1: The Impact of Various Degrees of Data Siloing Regulation ($\rho > 0$) and Data Sharing Regulation ($\delta > 0$) in Comparison to No Regulation ($\rho = \delta = 0$).

Note: Figures derived for $\theta = 1/10$.

Second, we could also allow competition in the primary market with platform 1 being the dominant inter-market player. This would not change our insights qualitatively, because it does not fundamentally change the mechanisms whereby data sharing or data siloing lead to lower data externalities in market B, and thus lower the innovation incentives of platform 1 in market A. However, the degree of competition in market A would moderate the impact of the regulation on innovation incentives in market A. The stronger the competition in A, the less would platform 1 reduce its level of innovation following a regulation. The additional insight from this extension is that competition in market A implies that the rival in market A innovates more with regulation and thus any negative impact of data sharing or siloing on welfare results is dampened due to competition in market A. However, one should also keep in mind that economic regulation can only be imposed, if a firm possesses significant market power, which we proxy here through a monopoly position in market A. Thus, in practice a scenario in which platform 1 faces significant competition in its primary market would most certainly preclude the possibility to impose data sharing regulation or data siloing regulation in the first place.¹⁰

Finally, future research could also seek to develop a more dynamic framework in which the domino effect and the successive entry of a dominant platform in more and more related markets could be studied. Our static, two-stage two-market model could provide a useful starting point for this.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number 409474477.

¹⁰ For example, under the current Digital Markets Act proposal, firms need to be a “gatekeeper” for a given “core platform service” (e.g. online search) and need to enjoy an “entrenched and durable position” in order to be subjected to any regulation (see Article 3 of the DMA). These criteria are not met if there exists effective competition with a rival platform service.

References

- Afuah, A. 2013. “Are network effects really all about size? The role of structure and conduct,” *Strategic Management Journal* (34:3), pp. 257–273.
- Agrawal, A., Gans, J., and Goldfarb, A. 2018. *Prediction machines: the simple economics of artificial intelligence*, Harvard, MA: Harvard Business Review Press.
- Argenton, C., and Prüfer, J. 2012. “Search engine competition with network externalities,” *Journal of Competition Law and Economics* (8:1), pp. 73–105.
- Bajari, P., Chernozhukov, V., Hortaçsu, A., and Suzuki, J. 2019. “The impact of big data on firm performance: An empirical investigation,” *AEA Papers and Proceedings* (109), pp. 33-37.
- Bundeskartellamt 2019. “Bundeskartellamt untersagt Facebook die Zusammenführung von Nutzerdaten aus verschiedenen Quellen,” Available at <https://www.reuters.com/article/us-tech-antitrust-google-quotebox-idUSKBN2752BU>. Last accessed 4/20/2021.
- Cennamo, C., and Sokol, D. D. 2021. “Can the EU Regulate Platforms Without Stifling Innovation?” *Harvard Business Review Digital Articles*.
- CMA 2020. “Online platforms and digital advertising final report. Appendix S: The relationship between large digital platforms and publishers.” Available at: https://assets.publishing.service.gov.uk/media/5efb22fbd3bf7f768fdcdfae/Appendix_S_the_relationship_between_large_digital_platforms_and_publishers.pdf. Last accessed on 08.04.2021.
- European Commission 2014. “Commission Decision of 03/10/2014 declaring a concentration to be compatible with the common market (Case No COMP/M.7217 - FACEBOOK / WHATSAPP) according to Council Regulation (EC) No 139/2004,” European Commission Communication, Retrieved from <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32014M7217>. Last accessed on 05/01/2021.
- European Commission 2017. “Mergers: Commission fines Facebook €110 million for providing misleading information about WhatsApp takeover,” Retrieved from https://ec.europa.eu/commission/presscorner/detail/en/IP_17_1369, Last accessed on 05/01/2021.
- European Commission 2020. “Digital Markets Act: Ensuring fair and open digital market,” Available at <https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:842:FIN>. Last accessed 4/19/2021.
- Gregory, R. W., Henfridsson, O., Kaganer, E., and Kyriakou, H. 2021. “The role of artificial intelligence and data network effects for creating user value,” *Academy of Management Review* (46:3), pp. 534-551.
- Haftor, D. M., Climent, R. C., and Lundström, J. E. 2021. “How machine learning activates data network effects in business models: Theory advancement through an industrial case of promoting ecological sustainability,” *Journal of Business Research* (131), pp. 196–205.
- Hagiu, A., and Wright, J. 2020. “Data-enabled learning, network effects and competitive advantage,” Working paper. Available at: <https://ap5.fas.nus.edu.sg/fass/ecsjkdw/data%20enabled%20learning%20june2020.pdf>.
- Humer, C., and Vengattil, M. 2020. “Alphabet’s Verily targets employer health insurance with Swiss Re partnership,” Available at <https://www.reuters.com/article/us-alphabetverily-idUSKBN25L1Q9>. Last accessed 2/26/2021.
- Katz, M. L., and Shapiro, C. 1985. “Network externalities, competition, and compatibility,” *The American Economic Review* (75:3), pp. 424–440.
- Krämer, J., and Schnurr, D. 2021. “Big Data and Digital Markets Contestability: Theory of Harm and Data Access Remedies,” *Journal of Competition Law and Economics* (<https://doi.org/10.1093/joclec/nhab015>)
- Parker, G., Petropoulos, G., and Van Alstyne, M. W. 2021. “Platform mergers and antitrust,” Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3763513.
- Prüfer, J., and Schottmüller, C. 2020. “Competing with big data,” *Journal of Industrial Economics* (forthcoming).
- Schaefer, M., Sapi, G., and Lorincz, S. 2018. “The effect of big data on recommendation quality: The example of internet search,” *DIW Berlin Discussion Paper* (1730). Available at: https://www.diw.de/de/diw_01.c.581649.de/publikationen/diskussionspapiere/2018_1730/the_effect_of_big_data_on_recommendation_quality_the_example_of_internet_search.html
- Shankar, V., and Bayus, B. L. 2003. “Network effects and competition: An empirical analysis of the home video game industry,” *Strategic Management Journal* (24:4), pp. 375–384.

- Tucker, C. 2019. “Digital data, platforms and the usual [antitrust] suspects: Network effects, switching costs, essential facility,” *Review of Industrial Organization* (54:4), pp. 683–694.
- Varian, H. R. 2014. “Beyond big data,” *Business Economics* (49:1), pp. 27–31.
- Vincent, J. 2019. “Facebook ordered to stop combining WhatsApp and Instagram data without consent in Germany,” Available at <https://www.theverge.com/2019/2/7/18215143/facebook-whatsapp-instagram-third-party-user-data-combined-banned-germanyfco-competition>. Last accessed 05/01/2021.
- Zhu, F., and Iansiti, M. 2012. “Entry into platform-based markets,” *Strategic Management Journal* (33:1), pp. 88–106.