



This article forms part of the Special Issue on The Governance of AI

Governance of data sharing: A law & economics proposal

Inge Graef^a, Jens Prüfer^{b,*}

^a LTMS, TILEC & TILT, Tilburg University, P.O. Box 90153, Tilburg 5000 LE, the Netherlands

^b Center, TILEC, Tilburg University, P.O. Box 90153, Tilburg 5000 LE, the Netherlands



ARTICLE INFO

Keywords:

Data sharing
Data-driven markets
Economic governance
Competition law
Data protection
Regulation

ABSTRACT

To prevent market tipping, which inhibits innovation, there is an urgent need to mandate sharing of user information in data-driven markets. Existing legal mechanisms to impose data sharing under EU competition law and data portability under the GDPR are not sufficient to tackle this problem. Mandated data sharing requires the design of a governance structure that combines elements of economically efficient centralization with legally necessary decentralization. We identify three feasible options. One is to centralize investigations and enforcement in a European Data Sharing Agency (EDSA), while decision-making power lies with National Competition Authorities in a Board of Supervisors. The second option is to set up a Data Sharing Cooperation Network coordinated through a European Data Sharing Board, with the National Competition Authority best placed to run the investigation adjudicating and enforcing the mandatory data-sharing decision across the EU. A third option is to mix both governance structures and to task national authorities to investigate and adjudicate and the EU-level EDSA with enforcement of data sharing.

1. Introduction

In the past two decades, the rate of technological progress has accelerated. Most of it has occurred in fields that draw heavily on machine-generated data about user behavior (Brynjolfsson and McAfee, 2012). Mayer-Schönberger and Cukier (2013) coined this development “the rise of big data” or “datafication,” which depends on two simultaneous technological innovations: first, the increasing availability of data, owing to improvements in information and communication technologies which easily and inexpensively store information such transactions produce or transmit (“big data”); second, the increasing ability of firms and governments to analyze novel big data sets, aided especially by machine-learning (ML) techniques (“artificial intelligence”, AI).¹

Many big data are generated while individual users interact with websites, apps, or programs (henceforth: services) of companies, who automatically log users’ choices and digital characteristics, e.g. their IP-

addresses and preferred languages. The more of such information service providers have, the better they can predict the preferences and other characteristics both of aggregate users and of individuals over time. We call such data *user information*.² Whereas many insights that service providers can infer from user information are hard to disentangle from the firm’s intangible assets, for instance, knowledge about the best ML model to draw value from user information, the raw user information can be shared relatively easily.³ This has important consequences for the distribution of value generated from these data.

Under the General Data Protection Regulation (GDPR), users have the right to receive personal data they have provided to a service provider and port it to another provider.⁴ However, depending on the specific market environment as well as the costs and benefits for the individual, users may often not have enough incentives to invoke their data portability rights in order to redistribute the value of data among providers. By contrast, on *data-driven markets* firms are highly incentivized to collect their

This article forms part of the Special Issue on The Governance of AI

* Corresponding author.

E-mail addresses: I.Graef@tilburguniversity.edu (I. Graef), j.prufer@uvt.nl (J. Prüfer).

¹ “AI is a bigger concept to create intelligent machines that can simulate human thinking capability and behavior, whereas, machine learning is an application or subset of AI that allows machines to learn from data without being programmed explicitly” (<https://www.javatpoint.com/difference-between-artificial-intelligence-and-machine-learning>). In this paper we use the terms interchangeably, referring especially to the ability of algorithms to learn and to develop themselves with very little human intervention.

² Argenton and Prüfer (2012) developed this definition of user information for the first time.

³ For example, search engines automatically save users’ queries and clicks on displayed lists of URLs in so-called *search logs* or *query logs*. These files can be shared in a standardized format.

⁴ Article 20 of Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (GDPR) [2016] OJ L119/1.

<https://doi.org/10.1016/j.respol.2021.104330>

Received 13 August 2020; Received in revised form 14 April 2021; Accepted 20 July 2021

Available online 6 August 2021

0048-7333/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

competitors' user information.⁵ Important examples include search engines, digital maps, platform markets (e.g. for hotels, transportation, dating, music/video-on-demand); probably also smart meters, self-driving vehicles, and various other industries.⁶

In this paper, we design a governance structure for the sharing of user information on data-driven markets and study how it complements existing regulatory tools to tackle monopolization of data-driven markets. To do so, we first draw on an economic governance framework⁷ in order to identify possible institutional structures of data sharing, which are centered around new organizations, coined the *European Data Sharing Agency (EDSA)* and the *European Data Sharing Board (EDSB)*. Then we characterize the optimal organizational governance structure, trying to minimize moral hazard of the involved decision makers. We also compare the proposed data-sharing governance scheme with alternative regulatory options to draw lessons for its design.

Due to the magnitude and significance of search engines and the other markets mentioned above for the entire economy, this exercise has value in itself. One of the main issues identified by the European Commission as obstacles for the EU to realize its potential in the data economy is the lack of available data for innovative re-use. To stimulate data exchange, the Commission has introduced a proposal for a Data Governance Act⁸ and is considering to introduce a Data Act by the end of 2021 as part of its European data strategy (European Commission, 2020a, p. 13). Although data-driven markets warrant a separate analysis of the optimal data-sharing governance structure, the economic governance methodology applied here can serve as a blueprint to study optimal data sharing in other contexts as well. It is also worth noting that the Commission in the proposed Digital Markets Act has introduced an obligation for search engine providers qualifying as gatekeepers (namely especially powerful providers meeting certain quantitative criteria)⁹ to provide third party providers of online search engines with access to ranking, query, click and view data on fair,

reasonable and non-discriminatory terms.¹⁰ While the scope of these legislative initiatives is—as we will argue – too limited to address the concerns in this paper, their introduction does illustrate the relevance of data sharing in current regulatory discussions.

2. Why and when mandate data sharing?

Over the last 15 years, an extensive literature on the economics of platforms and two-sided markets has developed, increasingly with a focus on competition policy problems. Julien and Sand-Zantmann (2021) summarize this literature and conclude that, besides software interoperability, the topic of data sharing needs more investigation. They interpret current policy moves and proposals to allow data portability as a consequence of these ideas but caution that the impact of giving more data access rights to consumers and competitors on major platform's business model is "far from being clear" (p.18). For the analysis of data sharing in the context of data-driven learning by doing, they refer to Hagi and Wright (2020) and Prüfer and Schottmüller (2017). See details below. Calvano and Polo (2021) confirm in their literature review that digital markets have a strong natural tendency towards concentration or market tipping, which suggests that models of competition for the market are more relevant than competition in the market.

In general, both the academic and the policy discussion about data sharing suffer from unclear definitions. For instance, most of the literature studies situations in which a consumer/user knows more about his/her type or willingness-to-pay for a service than the provider of the service.¹¹ The question is then, under which conditions the user is willing to (truthfully) share the private information, which is assumed to be costly, and what the welfare consequences of such sharing are. This is also relevant in a B-2-B context, where innovating firms may be less incentivized than socially efficient to share their insights or acquired data sets. Related are studies about "data commons" or "open data", which ask under which conditions governmental data should be shared with other parties.¹² All these approaches have in common that the starting point is asymmetric information and that the *voluntary* balancing of that information makes markets more efficient (or enables follow-on innovation) but comes at a cost for the individual, including a decrease in privacy, and, hence, the net welfare effects may be positive or negative.

By contrast, here we narrowly focus on *data-driven markets*, where the interaction between a service provider and a user is administered electronically such that it is possible to store users' choices (e.g. clicking behavior) and characteristics (e.g. location) with very little effort, i.e. virtually for free. Hence, the one provider who interacts with a user already has access to the user's data at the start of the analysis. Investigating data sharing in such an environment then asks what the consequences are if not only one provider, but many providers have access to user information. As we discuss in Section 3, this drastically reduces the amount of interactions to be studied as not millions of users have to give (costly) consent for data sharing but only very few providers have to share it. Moreover, in such an environment *mandatory* data sharing is needed because one party, the incumbent, has no incentives to share voluntarily.

Biglaiser et al. (2019) discuss the idea of data as a source of incumbency advantage (p.44): "Whether access to data can be considered an essential factor for competition in and for the market has been extensively debated in policy circles (e.g., see McAfee et al. (2015) and Varian (2015) for applications to search engines and Lambrecht and

⁵ See section 2 for details.

⁶ The development of an econometric test that establishes whether a certain industry is actually data driven and its application to one industry is under way (Klein et al., 2021).

⁷ See Dixit (2009) and Williamson (2005) for general introductions, Masten and Prüfer (2014) and Aldashev and Zanarone (2017) for game-theoretic models, and Prüfer (2013, 2018) for applications of this methodology to the problem of trust in cloud computing.

⁸ Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act), 25 November 2020, COM/2020/767 final. Note that the proposed Digital Services Act mandates a specific type of data sharing different from the data sharing between businesses considered in this paper, namely an obligation for platforms to share data with regulatory authorities and vetted researchers for the purpose of, respectively, monitoring compliance with the provisions of the Digital Services Act and identifying and understanding so-called systemic risks relating to the dissemination of illegal content, the exercise of fundamental rights and intentional manipulation of services offered by platforms with negative effects on public interests such as health and security. See Articles 31 and 26(1) of the proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act), 15 December 2020, COM (2020) 825 final.

⁹ Three main cumulative criteria apply for providers to be presumed a gatekeeper under Article 3(1) and (2) of the proposed Digital Markets Act: (1) a size that impacts the internal market: this is presumed to be the case if the company achieves an annual turnover in the European Economic Area equal to or above € 6.5 billion in the last three financial years, or where its average market capitalization or equivalent fair market value amounted to at least € 65 billion in the last financial year, and it provides a platform service in at least three Member States; (2) the control of an important gateway for business users towards final consumers: this is presumed to be the case if the company operates a platform service with more than 45 million monthly active end users established or located in the EU and more than 10 000 yearly active business users established in the EU in the last financial year; (3) an (expected) entrenched and durable position: this is presumed to be the case if the company met the other two criteria in each of the last three financial years.

¹⁰ Article 6(1)(j) of the proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act), 15 December 2020, COM(2020) 842 final.

¹¹ See Bergemann et al. (2020) and literature cited therein.

¹² See Yakowitz (2011) for an early example.

Tucker (2015) for a more general discussion).” In their literature review, they list Prüfer and Schottmüller (2017) as the only paper that, as of 2018, has proposed a specific model to capture the advantage of an incumbent/market leader in data-driven markets.¹³

Prüfer and Schottmüller (2021) define a *market as data driven* if a firm’s marginal costs of innovation decrease in the amount of user information, that is, if it is subject to specific feedback effects (“data-driven indirect network effects”).¹⁴ They show in a dynamic model of R&D competition that, in data-driven markets, user information leads to *market tipping (monopolization)*. The problem is that such a tipped market with one dominant firm and, potentially, a few very small niche players, is characterized by low incentives to innovate both for the dominant firm and for (potential) challengers.¹⁵

The intuition of this tipping tendency is that the smaller firms, even if they are equipped with a superior idea/production technology, face higher marginal costs of innovation because they lack access to the large pile of user information to which the dominant firm has access due to its significantly larger user base. Consequently, if a smaller firm were to heavily invest in innovation and roll out its high-quality product, the dominant firm could imitate it quickly-at lower cost of innovation-and regain its quality lead. The smaller firm would find itself again in the runners-up spot, which implies few users and low revenues-but it still has to pay the large cost for the attempted leap in innovation. Foreseeing this situation, entrepreneurs and private financiers would not invest in innovation of a smaller firm.¹⁶ In turn, because the dominant firm knows about the deterring disincentive to innovate for its would-be competitors, it is protected by its large (and constantly renewed) stream of user information and can rest on a lower level of innovative efforts, too. These *low innovation rates*, both by the dominant firm and by (would-be) competitors, as compared to a situation with lively competition, constitute the theory of harm in Prüfer and Schottmüller (2021).

Those authors also introduced the idea of *connected markets*: providers can connect markets if the user information they have gained is also valuable in another market. For instance, some search engine queries relate to geographic information. These data are also valuable when providing a customized map service. The authors showed that if the market entry cost in a “traditional” market are not too high, a firm that finds a “data-driven” business model can dominate any market in the long term. Relevant user information on its home market is a great facilitator for this process, which can occur repeatedly, generating a *domino effect*.

The third contribution of Prüfer and Schottmüller was, based on the earlier idea of Argenton and Prüfer (2012), to study the consequences of a regulatory intervention in data-driven markets: *to mandate the sharing of (anonymized) data on user preferences and characteristics amongst*

¹³ Martens (2021) offers a very accessible introduction to the economics of data and market power. He also acknowledges that data-driven network effects were first analyzed by Prüfer and Schottmüller (p.11).

¹⁴ Prüfer and Schottmüller (2021) is the published version of their 2017 working paper.

¹⁵ Complementing this research empirically, Schaefer et al. (2018) study with observational data from Yahoo.com whether there are economies of scale in internet search. They show that more data enhances search engine quality and that personal information (for instance, the ability of the search engine to track the browsing behavior of specific users) amplifies the speed of learning. Their findings are consistent with an incumbent data advantage due to possession of personal information. A similar result is shown by Bajari et al (2019) studying Amazon data. They find that the prediction accuracy of their models increases with the time dimension (but with diminishing returns to scale).

¹⁶ This result is reflected by Edelman (2015), who cites the oral testimony of Yelp’s CEO before the Senate Judiciary Subcommittee on Antitrust, Competition Policy and Consumer Rights on September 21, 2011, and writes: “Google dulls the incentive to enter affected sectors. Leaders of TripAdvisor and Yelp, among others, report that they would not have started their companies had Google engaged in behaviors that later became commonplace.”

competitors. They showed that, even in a dynamic model where competitors know that their innovation investments today affect their market shares and hence their innovation costs tomorrow, such a policy intervention could mitigate market tipping and would have positive net effects on innovation and welfare if data-driven indirect network effects are sufficiently strong.

Going one conceptual step further, Biglaiser et al. (2019) distinguish between *across-user learning* (e.g. in Google Maps), where the user information gained from one user helps a provider improve services for other users, and *within-user learning* (e.g. Google Nest), where one user’s consumption benefit from using the same service improves over time because the provider learns about his/her preferences and hence can better serve these. Hagiu and Wright (2020) put this distinction into a formal model and study the effects of data sharing. They show that, for across-user sharing, if a smaller firm anticipates such a policy, it may compete less aggressively in the first place, i.e. free-ride on the incumbent, which could potentially lower consumer surplus. However, even for this case Hagiu and Wright conclude (p.17): “While consumers may be better or worse off under data sharing, it is straightforward to confirm that [...] expected welfare is strictly higher with this type of data sharing whenever it implies a positive probability of [the entrant] winning, and is otherwise unchanged.” More importantly, in markets where nominal prices for consumers are already zero (as on markets for search engines or for online travel agencies), even this potential negative effect of data sharing does not apply. For within-user learning, Hagiu and Wright suggest that firms compete more aggressively in the absence of data sharing, which diminishes the potential benefits of data sharing.

Summarizing their results, Hagiu and Wright (2020:30) state, “we’ve shown that a key condition for data sharing policy to improve consumer surplus is that the firm that benefits from data sharing is sufficiently far behind the leader.” This view is echoed by our policy proposal, which holds that only firms with a market share of at least 30 % should be mandated to share their data. See section 3, Preliminaries, for details. This discussion is not only of academic value, but also feeds into policy debates where the adoption of measures to force access to privately held data is being considered (European Commission, 2020a).

In particular, in the 2019 report on “Competition Policy for the Digital Era” commissioned by EU Commissioner Vestager, Crémer et al. (2019:105/6) cite Prüfer and Schottmüller (2017) and then state: “The sharing of data with competitors may then promote competition and innovation in the industry, considering the non-rivalry of data use. [...] in these platform settings, another aspect may gain in relevance, namely the strong indirect network effects that such platforms-and in particular dominant ad-funded platforms-seem to be able to generate through their superior ability to monetise data. [...] Given [...] the data-driven feedback loops that tend to further entrench dominance, the benefits for competition and innovation to be expected from a mandated data sharing may then outweigh the negative effects on the dominant firm. In particular when it comes to access to data held by dominant platforms, there may, therefore, be a case for mandating data access.”

Based on a similar reasoning, the UK Competition and Markets Authority proposed in July 2020 that a new Digital Markets Unit should have the ability to order Google to share its click and query data with rival search engines to allow them to improve their algorithms (UK CMA, 2020:365/7). This was followed by the introduction by the European Commission of an obligation for gatekeepers offering search engine services to give third party search engine providers access to ranking, query, click and view data on fair, reasonable and non-discriminatory terms in its December 2020 proposal for a Digital Markets Act.¹⁷ The scope of our proposed form of data sharing is wider than this specific duty because it covers data-driven markets beyond search engines and targets undertakings beyond those qualifying as

¹⁷ Article 6(1)(j) of the proposed Digital Markets Act.

gatekeepers.¹⁸ The proposed Digital Markets Act also requires gatekeepers to provide business users free of charge with effective, high-quality, continuous and real-time access and use of aggregated or non-aggregated data that is provided for or generated in the context of the use of the platform by those business users.¹⁹ While this obligation can address issues relating to the dependence of business users on gatekeeping platforms, it does not mandate data sharing with rivals that is necessary to remedy the concerns we identify with regard to the overall competitiveness of data-driven markets.²⁰ For these reasons, the provisions of the proposed Digital Markets Act do not suffice to address our concerns.

The proposed Data Governance Act is relevant in this context, because it sets up a notification framework for so-called data sharing services that act as intermediaries for the exchange of data between businesses and between businesses and consumers.²¹ These data sharing services include platforms or databases enabling the exchange or joint exploitation of data, but also intermediation services between data subjects wishing to make their personal data available and potential data users when exercising the rights provided by GDPR.²² By keeping a register of providers of data sharing services and monitoring their compliance with several provisions in the proposed Data Governance Act, the EU legislator aims to increase trust in these services as additional mechanisms to stimulate voluntary data sharing. The proposed Data Governance Act also facilitates ‘data altruism’, where individuals or market players voluntarily make data available for the common good, by giving organizations the possibility to register as ‘Data Altruism Organisation recognised in the EU’.²³ Considering the voluntary nature of the data sharing facilitated by the proposed Data Governance Act,²⁴ the instrument cannot address our concerns either because these concerns relate to situations where market players have a strong commercial interest in keeping data to themselves.

3. Governance of data sharing on data-driven markets

We tried to explain the need for a mandatory data-sharing regulation on data-driven markets with references to the literature and identified the theory of Prüfer and Schottmüller (2021) as key input in the previous section. Klein et al. (2021) follow up on that theory and develop an econometric test that can identify empirically whether a market is data driven — and hence should be subject to mandatory data sharing-or not. As of this section, *we will assume that such a mandatory data-sharing obligation already exists* and develop three possible governance structures that can be implemented if the test has indicated that a certain industry is data driven.

3.1. Preliminaries: what and who?

What data should be shared on a data-driven market? As outlined above, only the sharing of user information is appropriate, no other data. These are raw data about users’ choices or characteristics, which can be logged automatically and at virtually zero marginal cost during a user’s interaction with a service provider. The policy proposal explicitly does not include processed data or even algorithms, in which the sharing party already invested effort, at positive marginal cost. If such data would be required to be shared, it might crowd out the dominant firm’s incentives to invest in analytics in the first place. This threat is not given if user information is shared as it is a free byproduct of the regular provider-user interaction. If only raw data are shared, it also incentivizes competitors to develop own models to analyze user information, which can lead to a plurality of approaches, differentiated products, and, hence, more choice for consumers.

Who should share data? Prüfer and Schottmüller (2021) propose that all firms active on a data-driven market could be obliged to share their user information in order to maximize the total amount of data available in the industry. This setup, however, neglects two factors. First, data sharing comes at a cost and creates an administrative burden (Jin and Wagman, 2021). Second, large firms are more likely to have access to other sources of information that complement user information from this market and hence have higher marginal benefits from user information received (especially regarding data used to train ML-algorithms).²⁵ As the goal of the policy proposal is, however, to establish a contestable level playing field, this suggests an *asymmetric data-sharing obligation*: large firms should share more data than small firms.

Moreover, the policy proposal does not only apply to markets that have already tipped but also to data-driven markets where a few firms still compete for dominance in the market.²⁶ Such races to the top, are characterized by very high incentives to innovate. In order not to distort them by only requiring one firm to share data, we propose the following rule: *Oblige a firm to share its user information if it has at least 30 % market share.*²⁷ A market share threshold of 30 % is also applied in Block Exemption Regulations under EU competition law. A safe harbor applies for instance to vertical agreements if the market share of each of the firms involved does not exceed 30 % and if the agreements do not contain certain types of severe restrictions of competition.²⁸ Under those conditions, the agreement is not considered to have competitive impact and falls outside the scope of the prohibition on collusion of Article 101 of the Treaty on the Functioning of the European Union (TFEU).

²⁵ Martens (2021:13) notes that “[mandatory data sharing] may reduce rather than increase competition when the data are hoovered up by large platforms that can offer users additional advantages, based on economies of scope in re-use and aggregation with other data sources.”

²⁶ For instance, in the search engine industry Google has more than 90% market share in all European countries, which makes search engines a tipped market par excellence (<https://gs.statcounter.com/search-engine-market-share/all/europe>). By contrast, among online travel agencies, potentially a data-driven market too, Booking and Expedia accumulated 41% and 32%, respectively of the industry’s global revenues (<https://www.statista.com/statistics/935028/revenue-distribution-of-leading-otas-worldwide/>).

²⁷ “Market share” in this sentence is to be understood broadly and flexibly. Ideally, it refers to the share of users or another proxy for the amount of user information collected. It could be used as a presumption, as is also the case for market share thresholds in EU competition law. The goal is that the 2nd/3rd largest firm’s operations are not stunted if it is no contender for the dominant-firm position. E.g. Tripadvisor follows Booking and Expedia but has only 4.6% of global revenues in the online travel agency market. They should not have to share data.

²⁸ Articles 2(1), 3(1), 4 and 5 of Commission Regulation (EU) No 330/2010 of 20 April 2010 on the application of Article 101(3) of the Treaty on the Functioning of the European Union to categories of vertical agreements and concerted practices [2010] OJ L 102/1.

¹⁸ See footnote 10 for the definition used in the proposed Digital Markets Act.

¹⁹ Article 6(1)(i) of the proposed Digital Markets Act.

²⁰ For completeness, it is worth mentioning that the proposed Digital Markets Act includes more obligations relating to the combination and use of data, such as Article 5(a) obliging gatekeepers to refrain from combining personal data across services unless the end-user has consented and Article 6(1)(a) obliging gatekeepers to refrain from using in competition with business users any data not publicly available that is generated through activities by those business users on the platform. However, these obligations are not targeted at the sharing of data with rivals as our paper focuses on and therefore are not sufficient to address the concerns we identify.

²¹ Articles 9–14 of the proposed Data Governance Act.

²² The data sharing services covered by the notification framework are listed in Article 9(1) of the proposed Data Governance Act.

²³ Article 15 of the proposed Data Governance Act.

²⁴ Note that the proposed Data Governance Act also introduces mechanisms to facilitate re-use of data held by public sector bodies in Articles 3–8. However, as we focus on data sharing by private companies these provisions do not address the concerns in this paper either.

Although our setting is different, this reasoning can be reversed such that a market share of 30 % indicates the ability of a firm to impact competition if it does not share its data. The presumption of dominance for a market share of 50 per cent under EU competition law²⁹ is not suitable, because it only captures the incumbent in the market. For the same reason, the introduction of an obligation to share search data in the proposed Digital Markets Act of the European Commission does not suffice to address our concerns, because the scope of the duty is restricted to a few large online platforms acting as gatekeepers.³⁰

Who should get the shared data? User information is a club good: it is nonrival and excludable. In a data-driven market, access to a sufficiently large amount of user information can be considered a necessary (not sufficient) condition to compete effectively. Therefore, it is efficient to share it with every party that can (potentially) use it as input into its own service and that benefits users in the end. Consequently, *user information should be shared with every organization that is active in the respective industry or that can explain how it would serve users with the data.* In line with the European Commission's plans to stimulate data sharing not only within but also across industries (European Commission, 2020a, p. 13), complementary uses of data in other industries or markets should also be enabled. However, we submit that incumbents qualifying as gatekeepers under the Digital Markets Act should not be entitled to receive data under the mandatory data-sharing regime. This would prevent that existing incumbents can take advantage of mandatory data sharing to get access to data from data-driven markets in which they are not yet active and expand their already strong competitive advantage with undesirable long-term risks as a result. The eligibility to receive data should apply independent of the receiving party's organizational form, that is, to for-profit, non-profit and public organizations (state authorities). In the same spirit, the appropriate access price to another provider's user information should be equal to the sharing provider's marginal cost of obtaining the user information, which is (roughly) zero.

3.2. An economic governance framework

Now we are in a position to apply the economic governance framework as sketched in Dixit (2009) and elaborated in Masten and Prüfer (2011, 2014). The first question we have to answer is what the economic governance problem is. Assume a data-sharing obligation for user information, as described above, is in place. Then, the two governance problems are: (i) whether the parties subject to the obligation (the top one to three companies in the industry) comply with it (in full); (ii) whether the receiving parties use the incoming data in a way that is in line with the spirit of the data-sharing obligation, namely to offer and improve services for end users while respecting several side constraints, especially privacy protection of the end users whose information is shared. If a party complies with its obligation, we say she *cooperates*. Otherwise, she *defects*.

For each transaction, an economic governance institution must identify (i) whether a player cooperated or defected (*adjudication*) and, in case of defection, (ii) who is supposed to take which action in order to punish the defector (*enforcement*). To tackle both issues, we apply Masten and Prüfer's (2011) classification of economic governance institutions, which is displayed in adjusted form in Fig. 1. This scheme lists and classifies potentially available institutions in general. The application to data sharing follows below.

The classification categorizes eight governance institutions (see the boxes in Fig. 1) that can potentially solve the identified problems. Beneath each pair of neighboring boxes, the short vertical line indicates a distinguishing characteristic of these two institutions, which applies to *all* institutions to the left or to the right of the vertical line (unless it is adjusted elsewhere). For instance, *social preferences* (= maximizing

utility of another person) are a characteristic only of *Internal Value Systems*, whereas *standard preferences* (= maximizing one's own utility) are a characteristic of all seven institutions to the right of it.

Regarding *enforcement*, only two fundamental technologies are available. First, *ostracism* or *boycott*: here some party threatens another one to cease the relationship with each other in case of defection, which would cost both parties the expected net present value from future cooperation. In Fig. 1, all institutions on the left, up to and including *arbitrators*, rely on ostracism to enforce behavioral rules of cooperation. The second enforcement technology is *coercion* (violence): here some party threatens a defector with punishment that has direct payoff consequences now. Coercion can occur immediately and, in contrast to ostracism, is not restricted to damage in the future and to the net present value of a specific relationship.³¹

Regarding *adjudication*, the classification ranges from a player's moral values on the very left of the scheme via private judgments by individuals (*bilateral interaction*) or decentralized groups (*social networks*), organized coordinators of such groups (in *associations* or *criminal organizations*) to publicly employed judges who are completely (*courts*) or partly (*regulators*) subject to written, general rules.

The classification is applied by asking whether a specific institution can solve the economic governance problems. It ranks institutions from left to right by increasing costs of using them. Hence, we start on the left. To cut the analysis short, we underline that the reliance on *ostracism* that connects bilateral interaction, social networks, and associations, strongly deteriorates the usefulness of these governance institutions for mandatory data sharing of user information on data-driven markets. There, a dominant firm would prefer, rather sooner than later, to cease sharing data with its competitors. In turn, the dominant firm has no interest in a long-term relationship with data receivers. Hence, it also has no substantive incentive to enforce cooperation on the data receivers' side by threatening them to stop delivering data if they breach end users' privacy because this would imply that the dominant firm has to monitor the operations of receivers, which is costly (and may furthermore infringe competition law).³² Hence, all institutions that rely on ostracism, so-called *private ordering* institutions (Bernstein et al., 2015) offer no solution to our problems.

The remaining three institutions all enforce rules (cooperation) via the threat of coercion. *Criminal organizations* are not suitable for our purpose because, by definition, they maximize someone's private objectives (usually profit or power maximization), whereas the goal of our entire exercise is to identify a solution to the economic governance problems that maximizes (consumer) welfare.

These considerations suggest that the solution to our problems lies in the realm of *public ordering*. Specifically, the combination of a presumed objective of consumer-welfare maximization with the enforcement powers of the state is critical. Only then, players can be expected to adhere to the desired rules such that firms with a data-sharing obligation share their relevant data completely and in a manner that is useful to receivers and receivers respect users' privacy rights.

This leaves two possible institutions on the very right of the classification scheme, regulators and generalist courts. *Generalist courts* are

²⁹ Case C-62/86, *AKZO*, ECLI:EU:C:1991:286, par. 60.

³⁰ Article 6(1)(j) of the proposed Digital Markets Act.

³¹ In practice, however, laws nearly always do not exploit the full potential of theoretically unrestricted punishment for rule transgressions. Violations of EU competition law are, for instance, capped at 10% of the total turnover of an undertaking.

³² Similarly, *internal value systems* are no solution. This institution refer to situations where an obliged party would not renege on her obligations because she is ethically motivated to act as promised. Ethics can install cooperative behavior in selected situations such as intra-family relationships, but given the strong incentives of the dominant firm to defect (see Prüfer and Schottmüller, 2021) and the increased value receiving parties could generate by defection, internal value systems offer no solution for implementing cooperation in data sharing

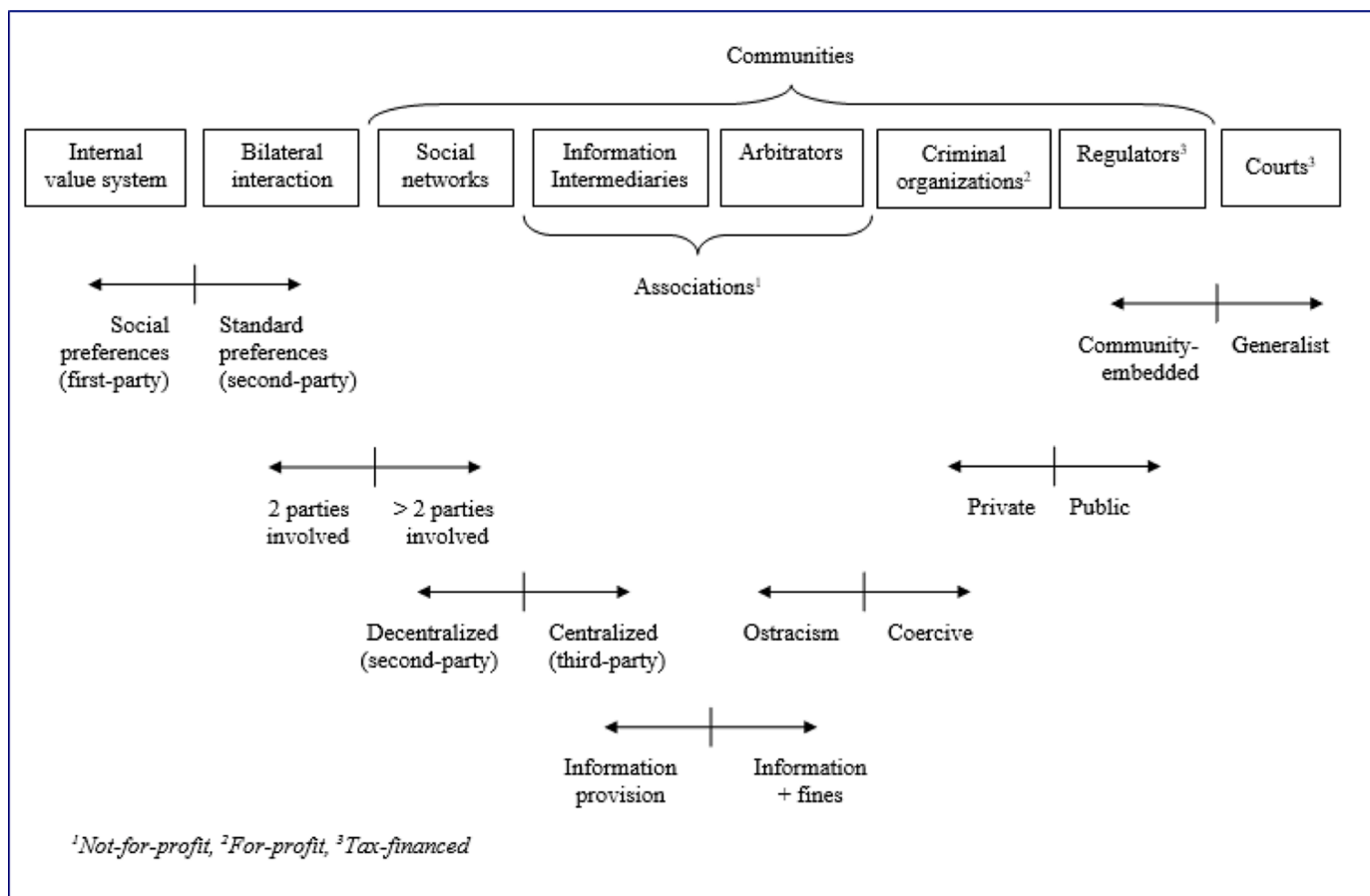


Fig. 1. A Classification of Economic Governance Institutions (adapted from Masten and Prüfer 2011).

bound by strict rules (laws) and staffed by judges who have extensive knowledge of those laws. Their main advantage is that they are as independent and impartial as is possible, from the institutional setup. Their disadvantage is, apart from being very costly to use,³³ that they have little knowledge of specific trades, for instance data-driven markets, and are not embedded in communities (see below).

Regulators improve these shortcomings of generalist courts. They operate with more flexible decision-making rules and are often staffed by experts in the subject from different disciplines (e.g. lawyers, economists, data scientists). Regulators are also embedded in (expert) communities, which implies that they not only understand a specific industry better than generalist judges, they also receive more information via informal channels and from a greater variety of sources about the relevant parameters of a given case. Consequently, the probability to err in a decision that requires expert knowledge is lower among regulators than in generalist courts.

The virtue that comes from community embedding of regulators is also their largest drawback: a person who has many friends in an industry can be easier influenced/corrupted/captured than a generalist judge, who decides plainly based on the law. We will minimize this problem via a specific governance structure in the subsection, “Organizational consequences of the governance options,” below.

Evaluating all trade-offs, we conclude that **regulators are the best available institution to solve the economic governance problems stemming from mandatory data sharing on data-driven markets.**

³³ The costs of using a governance institution comprise all costs, including those denominated in money, time, psychological stress, and other transaction costs.

Importantly, because the mandated data sharing we propose would become a new additional regulatory framework, it is not tied to an existing institutional or enforcement mechanism. Just like Regulation 1/2003 empowers competition authorities and courts to apply the EU competition rules,³⁴ the governance framework for data sharing would have to assign enforcement powers to certain institutions. Instead of starting from an *assumption* that regulators are the parties that should implement a certain legal provision based on experience in existing regulatory frameworks, we mapped all available options and regard the conclusion that regulators are the best placed institutions as a *result* of our analyze.

What “regulators” can mean in the practice of data-driven industries and how to mitigate the moral hazard/capture problem of those persons who decide about the details of data sharing and who could be lobbied by other industry participants is the subject of the next subsection.

3.3. Designing organizational governance: a European data sharing agency, a data sharing cooperation network, or both?

Above we established that, due to the opposed objectives of the firm (s) subject to the data-sharing obligation as compared to receivers of the data and users of their services, public ordering beats private ordering here. As a side-effect, our result of regulators as the best available institution implies that a *centralized* governance structure, where a third

³⁴ Council Regulation (EC) No 1/2003 of 16 December 2002 on the implementation of the rules on competition laid down in Articles 81 and 82 of the Treaty [2003] OJ L 1/1. For the approach in economics regarding institutions to govern data sharing, see all papers cited in Section 2.

party is in charge of organizing data sharing, is superior to a decentralized solution, where sharing and receiving parties are directly connected. The latter holds because, if n firms (up to three) have to share data and m firms (potentially hundreds) have a right to receive data, in a *decentralized* solution the number of direct connections is $n*m$, which can be large. This would increase the technological efforts (costs) necessary to manage the data exchange and also make it more difficult for any external party to monitor whether all firms cooperate. Consequently, the “regulator” in the *centralized* solution has to be a public agency that serves as *intermediary* between the n data-sharing firms and the m data-receiving firms. This changes the number of necessary links from $n*m$ to $n + m$.³⁵

The intermediary organization should be tasked with the structure and operation of the data-sharing scheme. It must have a legitimate mandate to perform its tasks in the entire EU as most data-driven markets are global, in the sense that user information gained in one local or regional or national market is also useful in another municipality/region/nation.³⁶ In case a relevant market is local/regional/national, e.g. platforms for food-delivery or services depending on a specific language, the intermediary organization should collaborate with national authorities in Member States.

Due to economies of scale and learning-curve effects, *ceteris paribus* it seems most efficient to create a new, independent EU body with fully centralized investigation and enforcement powers. However, the 1958 Meroni case law prohibits the delegation of discretionary powers to bodies not established by the EU Treaties,³⁷ unless these powers are precisely delineated and the margin of discretion is limited. As will be discussed below, an EU agency holding clearly defined competences as set by legislation would meet these requirements.³⁸ Another option is to embed the organization within the European Commission’s DG Competition. However, the competences of DG Competition are limited to the enforcement of the current EU competition rules.³⁹ As we discuss in [section 4](#), the governance framework for data sharing goes beyond the remit of existing EU competition law.

In terms of the legal basis for the mandatory data sharing we propose, two options are available. A first option is Article 352 TFEU, which can be used as a legal basis to adopt legislation for attaining one of the objectives set out in the EU Treaties where the Treaties have not otherwise provided the necessary powers. Protocol 27 on the internal market and competition, annexed to the EU Treaties, makes explicit that the EU can resort to Article 352 TFEU should it need new powers to protect competition in the internal market.⁴⁰ However, this legislative procedure does not enable the European Parliament to be involved as co-legislator and requires unanimity in the Council, which means that all Member States have to support the legislative proposal. A second-and in our view preferred option-is Article 114 TFEU, which is the legal basis for the adoption of legislation to harmonize national rules and to prevent regulatory fragmentation in the EU internal market. This provision is to be preferred because it does allow the European Parliament to act as co-

³⁵ The to-be-shared data would be collected from n firms, pooled centrally, and disseminated as a merged data set to m receivers. We will explain below how our result will effectively only require that n times data is reproduced. Cf. [Fig. 4](#).

³⁶ In principle, the scope of power should contain as many jurisdictions (and users) as possible. In practice, the highest level, where we can imagine that it is implemented, is the EU, for the time being.

³⁷ C-9/56, *Meroni*, ECLI:EU:C:1958:7 and C-10/56, *Meroni*, ECLI:EU:C:1958:8.

³⁸ Case C-270/12, *UK v. Parliament and Council*, ECLI:EU:C:2014:18, par. 53–54.

³⁹ As laid down by Article 105 TFEU. Article 103 TFEU forms the legal basis for the adoption of legislation to give effect to the principles set out in Articles 101 and 102 TFEU, but cannot be used here because the data sharing goes beyond the scope of the existing EU competition rules.

⁴⁰ Protocol No. 27 on the internal market and competition [2010] OJ C 83/309.

legislator and is satisfied with qualified majority voting in the Council. It is worth noting here that the proposed Digital Markets Act is also based on Article 114 TFEU.⁴¹ While the obligations in the proposed Digital Markets Act-as argued above-do not go far enough to address the problem of market tipping in data-driven markets, the mandatory data sharing we suggest could also be based on Article 114 TFEU. The European Commission is expected to publish a proposal for a Data Act in 2021 ([European Commission 2020a](#), p. 13) that may include additional duties to share data. In particular, the European Commission is prioritizing the development of so-called ‘European data spaces’ in certain sectors of key importance, such as manufacturing, agriculture, health, finance, energy and mobility ([European Commission 2020a](#), p. 22), which could potentially implement some of the ideas we outline here.

Any governance structure for data sharing must define who has control over three central tasks:

- 1 Investigation: Who collects and analyzes information about markets that could be data driven?
- 2 Decision making: Who decides whether a market is found to be data driven and, if so, who has to share which data and who can access those data?
- 3 Enforcement: Who sets up and manages the technological infrastructure necessary to share data and thereby monitors that mandatory data sharing is enforced properly?

Taking into account EU institutional limitations and drawing from existing experiences, we propose two alternative governance structures to implement data sharing and a mix of these two governance structures as a third option: first, the **creation of an agency at EU level** with autonomous investigation and enforcement powers under the control of the Member States; second, a **network of national authorities** with coordination at the EU level; and third, a **mixed option** consisting of a combination of involvement at the EU and national level.

3.3.1. Option 1: establishing a European data sharing agency

An example of an EU agency with enforcement powers is the European Securities and Markets Authority (ESMA) that is the single supervisor of credit rating agencies in the EU.⁴² The Court of Justice found that the prohibition on delegation of powers flowing from the Meroni case law does not apply to ESMA’s competences in the area of short selling, because they are precisely delineated and its margin of discretion is limited.⁴³ Translating these insights to our topic implies that competences to mandate data sharing can be delegated to a new EU agency, named the **European Data Sharing Agency (EDSA)** here ([Fig. 2](#)), as long as its competences are clearly defined and its margin of discretion is limited by conditions for intervention to be laid down in legislation.

While ESMA has its own investigation and enforcement powers, its decision-making power rests with the Member States within the Board of Supervisors that acts by simple or qualified majority.⁴⁴ The Board consists of a chairperson (non-voting), the heads of the national financial supervisory authorities (with voting power) and representatives from related agencies at the EU level (non-voting) as well as a representative

⁴¹ Article 1(1) of the proposed Digital Markets Act.

⁴² Regulation 1060/2009 of the European Parliament and of the Council of 16 September 2009 on credit rating agencies [2009] OJ L 302/1, as amended by Regulation 513/2011 and Regulation 462/2013.

⁴³ Case C-270/12, *UK v. Parliament and Council*, ECLI:EU:C:2014:18, par. 53–54.

⁴⁴ Article 44 of Regulation (EU) No 1095/2010 of the European Parliament and of the Council of 24 November 2010 establishing a European Supervisory Authority (European Securities and Markets Authority) [2010] OJ L 331/84, as amended by Directive 2011/61/EU and Directive 2014/51/EU.

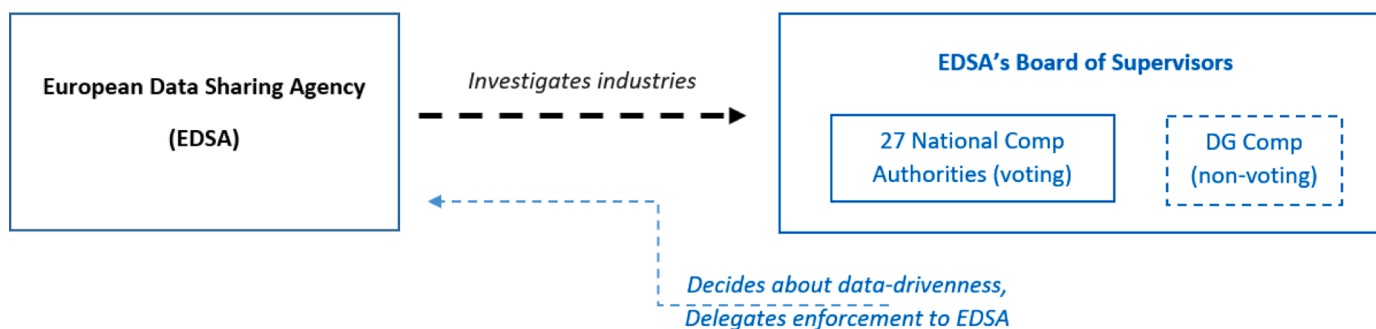


Fig. 2. Governance of data sharing [Option 1]: The European data sharing agency.

of the European Commission (non-voting).⁴⁵

Similarly, investigation and enforcement powers in the area of data sharing can be given to a newly established EDSA. Because of their experience in analyzing markets, national competition authorities (NCAs) are best placed to sit in a Board of Supervisors to be set up within EDSA together with a representative of DG Competition (non-voting, replicating ESMA's model). The advantage of the establishment of an EU agency is that the investigative tasks as well as the enforcement and technical implementation of data sharing are centralized at the EU level. The NCAs that are brought together in a *Board of Supervisors* take decisions based on the outcome of investigations conducted by EDSA at the EU level. The decisions of the Board of Supervisors to mandate data sharing are then again enforced by EDSA, within a data pool administered at the EU level by a technological unit to be set up within EDSA.

3.3.2. Option 2: Establishing a data sharing cooperation network

Based on the enforcement of EU data protection and consumer law,⁴⁶ the other option is to set up a network of national authorities coordinated at EU level, named the **Data Sharing Cooperation Network (DSCN)** here. Such a network implies decentralization of investigation and enforcement powers to Member State level. This seems costly but can also create opportunities for burden sharing across NCAs without having to establish a new specialized agency at EU level, partly duplicating investigation and enforcement powers already available at Member State level.

Both EU data protection and consumer law rely on a *one-stop-shop system*, where the outcome reached in a case with cross-border relevance investigated at the national level is effective in the entire EU. The enforcement approach within consumer protection, the so-called Consumer Protection Cooperation Network, to let the concerned authorities select the national consumer authority that is best placed to coordinate the case is more effective than the approach in the GDPR, where the national data protection authority of the main establishment of the respective firm is competent to act as lead supervisory authority.⁴⁷ The latter approach can be problematic if firms have their main establishment in Member States that do not have a strong data protection authority with enough resources to investigate cross-border cases.

A similar enforcement system can be designed for data sharing (Fig. 3): the Commission's **DG Competition takes the initiative** by notifying the (to be constructed) **DSCN**, in which the NCAs are organized, about markets that are potentially data-driven. An analogy can be made here with the EU electronic communications framework, whose enforcement also takes place at the national level but where the Commission sets out a recommendation containing the relevant markets

that are in its view susceptible to ex ante regulation at the national level.⁴⁸ The authorities involved jointly designate **one NCA that is best placed to lead the investigation, the so-called Lead NCA**. The latter investigates the industry in question and prepares a draft decision on which the other authorities have an opportunity to comment in an endeavour to reach consensus.

In data protection law, the GDPR prescribes a *consistency mechanism* that lays down a procedure in case a data protection authority objects to the draft decision of the lead authority.⁴⁹ In such situations, the European Data Protection Board, consisting of the head of all national data protection authorities, the European Data Protection Supervisor and a representative of the EU Commission (non-voting),⁵⁰ becomes involved and can ultimately adopt a binding decision if the lead authority rejects objections raised by other authorities.⁵¹

Transferring this scheme to mandatory data sharing suggests that, if the relevant NCAs object to the decision proposed by the Lead NCA and the Lead NCA rejects the objections, the (to be created) **European Data Sharing Board (EDSB)** takes the final decision. The EDSB consists of the heads of all NCAs and a representative of DG Competition (non-voting). Even though this governance structure is more decentralized than the establishment of an EU agency, it is still relatively effective because there is **only one decision on the data-drivenness of markets and only one authority that implements the data-sharing obligation with an effect across the entire EU**.

3.3.3. Option 3: Mixing governance regime

These two governance options can also be mixed, such that the technical infrastructure to enforce data sharing is set up at the EU level within the EDSA, while the investigation and decision-making powers are delegated to NCAs in a DSCN. Table 1 summarizes the major governance options, where "EU" means that the respective task rests at the supra-national EU-level and "Nat" refers to the national Member State level.

3.3.4. Organizational consequences of the governance options

Option 2, to establish a DSCN as a network among national authorities, requires extending the organizations of all 27 NCAs in the EU with the capacity to administer data pools. Each NCA would need two organizational units, which have structurally different tasks and, hence, have different governance structures: an *investigative unit* and a *technological unit*. Option 1, to establish an EDSA as a novel EU agency, has the advantage that investigations and enforcement can take place at EU level with involvement of NCAs only at the decision-making stage, such that only one investigative and only one technological unit will have to

⁴⁵ Article 40 of Regulation (EU) No 1095/2010.

⁴⁶ Article 68(5) GDPR and Article 1 of Regulation (EU) 2017/2394 of the European Parliament and of the Council of 12 December 2017 on cooperation between national authorities responsible for the enforcement of consumer protection laws (CPC Regulation) [2017] OJ L 345/1.

⁴⁷ Compare Article 17(2) CPC Regulation and Article 56(1) GDPR.

⁴⁸ See Commission Recommendation of 9 October 2014 on relevant product and service markets within the electronic communications sector susceptible to ex ante regulation [2014] OJ 295/79.

⁴⁹ Article 60(4) and 63 GDPR.

⁵⁰ Article 68(3) and (5) GDPR.

⁵¹ Article 65(1)(a) GDPR.

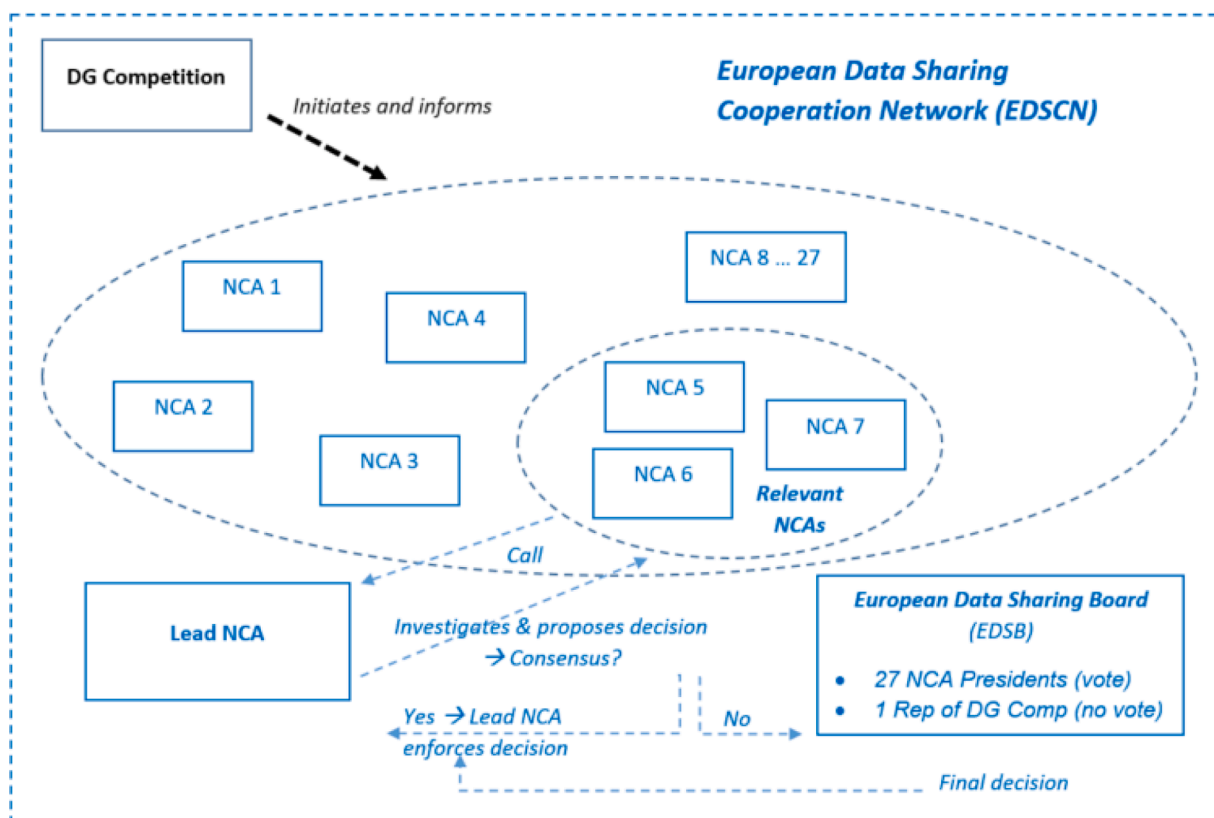


Fig. 3. Governance of data sharing [Option 2]: The data sharing cooperation network.

be created within the EDSA.

If the DSCN under option 2 calls a national authority to serve as Lead NCA for a potentially data-driven industry or the EDSA identifies one under option 1, its *investigative unit* serves as the face and the brain of the Lead NCA/EDSA, similar to today’s case teams in competition authorities. This unit has to conduct the test for data-drivenness mentioned above. Once a data-driven market is identified, the unit must determine which firms are subject to a data-sharing obligation and which firms have a right to access the shared data. Therefore, it must identify which data have to be shared and validate the business plans of potential market entrants (to check whether they qualify for receiving data). Under option 2, the Lead NCA prepares a draft decision and the consistency mechanism is applied if other NCA’s object to this decision. Under option 1, the investigative unit of EDSA prepares the decision while the Board of Supervisors, within which NCAs have voting power, decides on its adoption.

The *technological unit* serves as the hand and heart of the respective authorities. Its main task is to set up and run the technological infrastructure that ensures that data-receiving firms have access to the relevant data of data-sharing firms. This unit is either located within each of the NCAs in case of option 2 or within the EDSA in case of option 1.

The key feature of the “Mixed” governance regime under option 3 is that the two tasks of investigation and enforcement are separated, that is, the investigative and technological units working on one industry are not under the same organizational roof. As both tasks are decoupled in time (enforcement only starts after decision making, which follows investigation) and also have little overlap in their required expertise

(law and economics for *investigation*, computer science for *enforcement*), we do not regard this separation as problematic, though.

By contrast, this governance structure has three features that implement *checks and balances*. Together they ensure that the authority in charge of implementing mandatory data sharing, which serves as the “regulator” identified as optimal above, is *not captured* by partisan interests:

- 1 Separating tasks in an investigate and a technological unit.
- 2 While staff in the investigative unit should be civil servants, just as in today’s national or European competition authorities (and hence subject to orders from administrative superiors), there is no need to subject the technological unit to same hierarchy. Instead, it could be run by independent domain experts, just as the European Central Bank, who only have a technological task, fixed term limits, and restrictive cooling-off periods after leaving the technological unit to avoid switching jobs to regulated industry quickly.
- 3 The consistency mechanism: In governance options 2&3, if the Lead NCA was captured, other NCAs could object the decision proposal. In option 1, if the EDSA was captured, the Board of Supervisors/NCAs could object.

3.3.5. *Don’t share data, pool it and invite learning algorithms!*

Regardless of which option is implemented, a key challenge is to set up a scheme that protects the privacy of end users. Even if a large firm with access to other data sources collects personal information by interacting with a user, after sharing user information it must be

Table 1 Allocation of core data-sharing tasks across three governance options.

Tasks	Option 1: “EDSA”	Option 2: “DSCN”	Option 3: “Mixed”
Investigation	EDSA (EU)	Lead NCA (Nat)	Lead NCA (Nat)
Decision making	NCAs (Nat)	DSCN (Nat)	DSCN (Nat)
Enforcement	EDSA (EU)	Lead NCA (Nat)	EDSA (EU)

impossible to trace this information back to the individual user. Computer scientists have developed several different technologies to achieve this goal. Here, we sketch two promising concepts but leave the details to experts from that discipline.⁵²

- 1 *Anonymization (and synthetic data)*: Several ways of anonymizing data exist. The problem with many is that, if only some identifiers are removed from personal data (“pseudonymization”) and the data-receiving firm has access to other relevant data sources, it may be possible to re-identify individuals. In turn, if the shared data are reduced to aggregate information without any possibility to link it back to individuals, its value for data-receiving firms is sincerely diminished. This thwarts the original goal of the data-sharing obligation, to create a level playing field among competitors in a data-driven market.⁵³
- 2 *Data protection (and data pooling) behind a curtain*: One problem of anonymizing and sharing data is, as sketched above, the need to share large amounts of data via $n + m$ links. This may be technologically possible⁵⁴ but it will create the $(n + m)$ -fold multiplication of the original data sets, which offers other parties, including criminal hackers, many opportunities to access the shared data for unwarranted purposes. An alternative is to *not share* the user information with other firms but to *pool* it “behind a curtain”, managed by the Lead NCA/EDSA, and to offer firms with a right to “receiving” the shared data to send their ML-algorithms to the pool and let them be trained there.⁵⁵

Fig. 4 illustrates the organizational implementation of data sharing via a data pool.

The first advantage of this scheme is that data flows only via $n \leq 3$

⁵² See for instance the discussion of anonymous use of individual-level data in Cr mer, De Montjoye and Schweitzer (2019:85–87).

⁵³ One (imperfect) solution out of this dilemma is *synthetic data*. Here, an artificial (synthetic) data set is created that has the same aggregate characteristics as the original to-be-shared data set. However, as the shared data set is artificial, no real individuals can be re-identified. It seems that, with synthetic data, the value that can be derived from cross-section analyses of the original data set can be maintained. However, the time-series value, which stems from knowing what user X liked in the past when serving her in the present, cannot be transferred to receiving firms in this way. See Belovin et al. (2019) for a discussion, https://en.wikipedia.org/wiki/Synthetic_data for an introductory explanation, and <https://www.syntho.ai/> for an application. An alternative solution may consist of “data vaccination,” where content and personal identifiers are split and saved in different databases, which are only brought together again when an application is run. See <https://www.datavaccinator.com/>.

⁵⁴ Recently, Google, Facebook, Microsoft, Twitter, and other firms showed that sharing of user data is technically and organizationally possible at a large and automatic scale. They had announced a new standards initiative called the *Data Transfer Project*, designed as a new way to move data between platforms. See <https://www.theverge.com/2018/7/20/17589246/data-transfer-project-google-facebook-microsoft-twitter>.

⁵⁵ In personal conversation, a high-ranking computer scientist of a search engine company confirmed that this would be technically possible in his industry. The Big Data Value Association also proposed recently to allow data producers to retain their data locally and only allow specific algorithms (authorized “apps”) to perform approved functions locally without giving access to the raw data to anyone else (Vallejo et al., 2019). Findata, a service for the secondary use of health and social data in Finland operating since 2020, has a related concept (<https://www.findata.fi/en/>). There, however, interested parties can only ask a data permit authority to collect and analyze sensitive data on behalf of them. It is not possible to let (interested parties’) algorithms perform the work, which limits the scalability of the scheme and, hence, makes it impractical for the big data sets that have to be shared or accessed in data-driven markets. The concept of Data Safe Haven is closely connected (<https://www.ed.ac.uk/information-services/research-support/research-data-service/during/data-safe-haven/intro-data-safe-haven>).

links (instead of $n + m$ or even $n \cdot m$) and that only one data pool with the merged data of the n sharing firms exists, which reduces both the costs and the risks of data sharing. The second advantage is that, because the m competitors of the n sharing firms do not receive the raw data, privacy risks are reduced.⁵⁶ The m firms’ algorithms-and no human being-“see” the raw data but cannot take it outside of the data pool. Instead, they can only take the insights from their analyses outside: parts of the m firms’ services might even be serviced by algorithms operating from the data pool administered by the Lead NCA/EDSA. This alleviates the need to anonymize the data in the pool, which secures its full value for the m firms and their users and, hence, can contribute to establishing a level playing field in data-driven markets.

Given this massive amount of important tasks, the Lead NCA/EDSA must be well equipped with resources, especially with experts from various domains (mainly from law and economics in the investigative unit, and from computer science and data science in the technological unit). It also requires appropriate regulatory powers to perform its tasks effectively.

4. Insufficiency of existing regulatory options and legal constraints for the governance of data sharing

Section 3 contained our proposal for a governance structure of data sharing on data-driven markets. Now we explain what insights can be drawn from the existing regimes of EU competition, data protection and intellectual property law to implement the governance structure, beyond the elements already incorporated above. On the one hand, the limits of these regimes in facilitating data sharing show what new additional actions are needed for effective redress against market tipping. On the other hand, existing legal frameworks impose boundaries that a governance structure for data sharing will have to incorporate into its design. This includes obligations under data protection law to guarantee the privacy of individuals as well as the exclusivity offered under intellectual property law that will be under pressure if intellectual property protected data needs to be shared.

4.1. Competition law

Competition law is relevant to the issue of data sharing for two reasons. First, voluntary sharing of data among market players may give rise to *collusion*, and second, refusals to share data may amount to *abuse of dominance*. This situation may seem paradoxical but can be explained by the different scope and purpose of the prohibitions on collusion and abuse of dominance.

4.1.1. Data sharing as collusion

The prohibition on collusion of Article 101 TFEU protects against harm arising from agreements or concerted practices between market players that restrict competition. Data sharing arrangements (also referred to as ‘data pooling’) will often be procompetitive, because they lower entry barriers to the market and increase consumer choice for products and services (Lundqvist, 2018). However, such arrangements may also result in restrictions of competition when data sharing enables competitors to become aware of each other’s market strategies or when access to a data sharing arrangement is limited to certain market players (European Commission, 2019). The governance structure for data sharing proposed here involves the exchange of raw user information and not information further processed by firms, so that the system is unlikely to facilitate the sharing of commercially sensitive information. In addition, by relying on a centralized authority to organize the data

⁵⁶ While the privacy risks are reduced, it needs to be recognized that machine-learning models can allow for inferences to be made about characteristics of the training data and can even be considered personal data in themselves. See Veale et al. (2019).

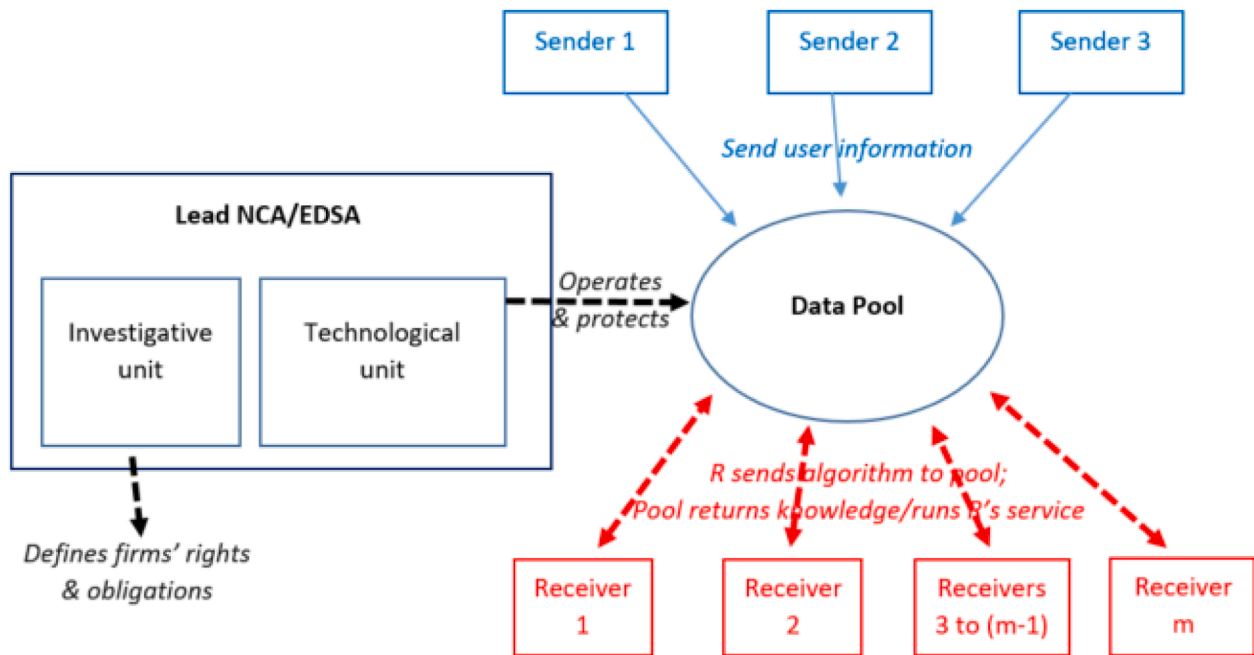


Fig. 4. Implementation of data sharing on data-driven markets via a data pool.

sharing, competitive risks can be constantly monitored.

4.1.2. Refusals to share data as abuse of dominance

The prohibition on abuse of dominance of Article 102 TFEU protects against abusive conduct by dominant firms excluding competitors or exploiting consumers. Data sharing can be imposed as a remedy to address a competition problem in a market. Most relevant for our purposes are cases where a dominant firm refuses to share data upon a request from a third party. Refusals to deal, including refusals to share data, can amount to abuse of dominance under Article 102 TFEU. To test the anti-competitiveness of a refusal to deal, the so-called *essential facilities doctrine* is relevant. This doctrine lays down the conditions under which a refusal to grant access to an essential input held by a dominant firm is abusive. Because of the far-reaching impact of the imposition of a duty to deal on the freedom to contract and the right to property of the dominant firm, the EU Courts have established a high legal hurdle for a refusal to deal to violate Article 102 TFEU. Only in the presence of 'exceptional circumstances' a refusal to deal gives rise to abuse of dominance, namely when: (1) the input is indispensable, (2) effective competition on a downstream market is excluded, (3) the emergence of a new product is prevented (this condition only applies to intellectual property protected assets), and (4) an objective justification is absent.⁵⁷

At the national level, competition interventions have taken place to share datasets in order to open up more traditional markets, including those for the supply of gas⁵⁸ and lottery services.⁵⁹ To our knowledge, no cases regarding abuse of dominance have so far occurred at the EU level

to force access to data in the data-driven markets we focus on here.⁶⁰ Competition policy reports do point at the need to consider mandating data access in particular in the context of the platform economy where due to the economic characteristics 'the benefits for competition and innovation to be expected from a mandated data sharing may then outweigh the negative effects on the dominant firm' (Crémer, De Montjoye and Schweitzer, 2019, p. 106). While previous cases met the high legal threshold under which a refusal to deal amounts to abuse of dominance,⁶¹ the tendency of data-driven markets to tip is a structural problem and therefore requires a remedy whose scope goes beyond the exceptional circumstances of the current essential facilities doctrine. The introduction by the European Commission of an obligation for gatekeepers offering search engine services to give third party search engine providers access to search data in its proposed Digital Markets Act does not provide a structural solution either, considering that the scope of the duty is restricted to search data and to a few large online platforms acting as gatekeepers.⁶²

4.1.3. Data protection law

The GDPR contains the relevant legal rules applicable to the processing of personal data defined as 'any information relating to an identified or identifiable natural person (data subject)'.⁶³ Although the GDPR does not prohibit the processing of personal data as such, some of the conditions it imposes restrict the sharing of personal data. At the same time, the GDPR has introduced a new *right to data portability* for

⁵⁷ Joined cases C-241/91 and C-242/91, *Magill*, ECLI:EU:C:1995:98; Case C-7/97, *Oscar Bronner*, ECLI:EU:C:1998:569; Case C-418/01, *IMS Health*, ECLI:EU:C:2004:257; Case T-201/04, *Microsoft*, ECLI:EU:T:2007:289. With regard to refusals to license intellectual property rights, the General Court has accepted that a limitation of 'technical development' is sufficient to satisfy the new product requirement. See Case T-201/04, *Microsoft*, ECLI:EU:T:2007:289, par. 647.

⁵⁸ French Autorité de la concurrence, Decision n°17-D-06 (*GDF Suez*), 21 March 2017, available at <http://www.autoritedelaconcurrence.fr/pdf/avis/17d06.pdf>.

⁵⁹ Decision 2015-P/K-27 of 22 September 2015 of the Belgian Competition Authority, *Stanleybet Belgium/Stanley International Betting and Sagevas/World Football Association/Samenwerkende Nevenmaatschappij Belgische PMU v. Nationale Loterij*.

⁶⁰ Remedies relating to access to data have been adopted in EU merger cases, such as Case No COMP/M.4726– *Thomson Corporation/Reuters Group*, 19 February 2008 and Case M.8124–*Microsoft/LinkedIn*, 6 December 2016.

⁶¹ Some of these cases can be interpreted as relating to information assets more broadly, such as the refusal to license a copyrighted brick structure for data on regional sales of pharmaceutical products in *IMS Health* and the refusal to share interoperability information needed for rivals to build software for the Windows operating system in *Microsoft*.

⁶² Article 6(1)(j) of the proposed Digital Markets Act.

⁶³ Article 4(1) GDPR: 'an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person'.

data subjects, which facilitates the exchange and re-use of personal data.⁶⁴ The scope of these obligations for the processing of personal data and the right to data portability are important to understand the relationship of data protection law towards data sharing.

4.1.4. Compatibility of data sharing with the GDPR

One of the key conditions laid down by the GDPR is the need to have a lawful ground for the processing of personal data.⁶⁵ Possible lawful grounds include consent of the data subject, performance of a contract, compliance with a legal obligation, and legitimate interests of the data controller.⁶⁶ Sharing of personal data is form of processing and therefore requires a lawful ground. While there are possibilities to anonymize personal data before sharing, rely on synthetic data or conduct data pooling behind a curtain as discussed in section 3, the usefulness and technical feasibility of these options is not yet completely clear. Despite their promising nature to prevent data protection issues from occurring, risks remain so that a governance structure for data sharing should incorporate data protection compliance into its design. This means that the **data sharing needs to have a lawful ground and must comply with the other obligations of the GDPR.**

A *duty to share* personal data as mandated by a newly established legal regime for data sharing can in principle be based on the lawful ground of compliance with a legal obligation, just like the imposition of a *duty to deal* under competition law can constitute a legal obligation to share personal data (Graef, 2016; Tombal, 2020; but arguing differently, see Kathuria and Globocnik (2020)). This would imply that consent of the individuals whose personal data will be shared on the basis of a new law is *not* needed. While such a reading is correct from a strictly legal perspective, it will be controversial to require firms to share personal data with third parties without requiring any affirmative action on the part of the individuals affected. Sufficient safeguards should be integrated into the design of the data sharing regime to eliminate data protection risks. In particular, the personal data to be shared and its further use should be as confined as possible. The principles of purpose limitation⁶⁷ and data minimisation⁶⁸ are relevant as restrictions here.

The principle of *purpose limitation* entails that personal data can only be collected for specific purposes and cannot be further processed in a manner that is incompatible with those purposes. In the context of data sharing, this implies that the shared personal data can only be processed in line with the purpose for which the data was shared as set out in new data sharing legislation. If the shared personal data is further processed in a way that does not fit with this purpose, a separate lawful ground is again required unless the new purpose can be considered as compatible with the original purpose.⁶⁹ An example of an incompatible purpose would be when the new provider also starts using the shared personal data to engage in direct marketing, while the legislation mandated data sharing only to improve search results.

The *data minimisation* principle requires personal data to be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed. For data sharing, this implies that the amount of personal data to be shared should be as limited as possible for the purpose at hand. The advantage of the centralized nature of the governance structure for data sharing proposed here, around the Lead NCA/EDSA, is that an overall procedure can be set up that would apply to all transactions between firms and that is administered by an authority. Such a centralized procedure reduces data protection risks, as compared to a decentralized model where such risks would have to be

addressed bilaterally between parties. In addition, because a regulator serves as intermediary between n data-sharing firms and m data-receiving firms, the number of necessary links is reduced from $n \cdot m$ to $n + m$ and even to $n \leq 3$ links when data pooling occurs behind a curtain as explained in section 3. At the same time, when personal data is pooled by the authority itself, the security of its system should be a key point of attention.

4.1.5. Why data portability does not suffice

While the GDPR thus imposes conditions for data sharing through obligations for data controllers, it has also introduced a right to data portability that facilitates sharing and re-use of personal data at the request of the data subject. Article 20(1) GDPR gives data subjects a right 'to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format' and transmit those data to another controller 'without hindrance'. Article 20(2) GDPR provides data subjects with the right to have personal data transmitted directly from one controller to another 'where technically feasible'. The exact scope of many of these terms remains unclear and is likely to be clarified through future litigation.⁷⁰

Irrespective of how these terms will be interpreted in the future, the GDPR's right to data portability is unlikely to be capable of addressing the risk of market tipping in data-driven markets. The porting of data creates a positive externality through the better predictions, search results etc. that all users will receive when an additional user brings her personal data to a new provider. However, as users typically do not take this benefit into account when deciding to port data, we should expect too little data portability requests to remedy market tipping (Graef and Prüfer 2018, p. 300). In addition, the GDPR's right to data portability does not require data controllers to delete ported personal data. As a result, a dominant firm does not lose data and will thus still own more data than rivals-even if those rivals can now access parts of that data following individual data portability requests. When invoking their right to data portability, data subjects may simultaneously invoke the right to erasure but there is no complete alignment between the scope of application of these independent rights (Graef et al., 2018). It is worth noting that the proposed Digital Markets Act includes provisions facilitating real-time, continuous portability for businesses and end-users.⁷¹ While these provisions expand the GDPR's right to data portability, they still do not form a sufficient solution for the problem of market tipping as rivals remain dependent on whether businesses and end-users invoke data portability.

4.2. Intellectual property law

There is no specific property right for data as such, but data may fall under the scope of protection of existing intellectual property rights. This includes *copyright*, *sui generis database protection* and *trade secret protection* (Drexler, 2017). Copyright law protects the original expression of an idea and grants authors temporary exclusive rights.⁷² While facts or data in itself do not qualify for copyright protection, databases

⁷⁰ See for instance Article 29 Working Party, Guidelines on the right to data portability, 5 April 2017, WP 242 rev.01, interpreting the notion of 'provided' personal data and the stipulation of Article 20(4) GDPR that the application of the right to data portability should not adversely affect the rights and freedoms of others (including the data protection rights of others as well as the intellectual property rights held by data controllers).

⁷¹ See Article 6(1)(h) and (i) of the proposed Digital Markets Act.

⁷² Articles 2 and 3 of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (Information Society Directive) [2001] OJ L 167/10 as amended by Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market [2019] OJ L 130/92.

⁶⁴ Article 20 GDPR.

⁶⁵ Article 5(1)(a) GDPR.

⁶⁶ Article 6(1) GDPR.

⁶⁷ Article 5(1)(b) GDPR

⁶⁸ Article 5(1)(c) GDPR

⁶⁹ Recital 50 GDPR. See the requirements to assess compatibility in Article 6(4) GDPR.

created by providers based on information they have collected may be copyright-protected. Databases can be protected either through copyright as regards the structure of original databases or through the sui generis database right as regards the contents of databases in so far as a substantial investment has been made in either the obtaining, verification or presentation of these contents.⁷³ In addition, data or information may benefit from trade secret protection if it is secret, has commercial value because it is secret and has been subject to reasonable steps to keep it secret.⁷⁴ The difference with copyright and sui generis database protection is that trade secret protection does not grant right holders an exclusive right to prevent third parties from using the subject matter of protection, but only protects against unlawful acquisition, use and disclosure.⁷⁵ In other words, trade secret protection cannot be invoked against the use of information obtained through legitimate means such as independent discovery or creation.⁷⁶

The extent to which data is protected by intellectual property rights or trade secrets is relevant for the issue of data sharing, because the creation of new duties to share data will limit the scope of protection offered to right holders. By obliging right holders to share data, their right to exclude third parties from using the protected subject matter (namely the data) and the secrecy of information is taken away. The invasion into intellectual property rights can be made proportional by incorporating such considerations into the design of new measures for data sharing. Furthermore, intellectual property protection is not absolute and can be balanced against other rights or interests. For instance, refusals to license intellectual property rights have been held abusive under EU competition law, resulting into duties to license and thus limiting the exclusivity provided to intellectual property right holders.⁷⁷

A similar balancing of interests can be applied here. The governance structure for data sharing we propose mandates the *exchange of raw user information only*. The data to be shared does not include insights gained about users resulting from further data processing or investments. This implies that tensions with intellectual property protection are limited due to the very design of the data sharing measures.

Two main insights can be derived from this section. On the one hand, **the analysis of the available mechanisms under competition and data protection law** (namely, competition law's essential facilities doctrine and the GDPR's right to data portability) **indicates that they are insufficient to solve the issue of market tipping** discussed in section 2. On the other hand, **the governance structures proposed in section 3 already incorporate the limits established by data protection and intellectual property law** (namely, the GDPR's requirements for a lawful ground for data processing, purpose limitation and data minimization and a balancing with the exclusivity offered by intellectual property rights). This means that no legal constraints would stand in the way of the mandatory data sharing we propose. We recommend for such new data sharing duties to be introduced in a regulatory framework complementing the existing possibilities to facilitate data sharing under competition law and data protection law. In other words, competition law's essential facilities doctrine and the

⁷³ Article 3(1) and 7(1) of Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L 77/20.

⁷⁴ Article 2(1) of Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure (Trade Secrets Directive) [2016] OJ L 157/1.

⁷⁵ Subsections 2–5 of Article 4 of the Trade Secrets Directive specify the circumstances in which the acquisition, use and disclosure of a trade secret is considered unlawful.

⁷⁶ Article 3 of the Trade Secrets Directive lists situations in which the acquisition, use and disclosure of a trade secret is considered lawful.

⁷⁷ See Joined cases C-241/91 and C-242/91, *Magill*, ECLI:EU:C:1995:98; Case C-418/01, *IMS Health*, ECLI:EU:C:2004:257; and Case T-201/04, *Microsoft*, ECLI:EU:T:2007:289.

GDPR's right to data portability would still apply in parallel to capture scenarios not addressed by the newly established duties to share data.⁷⁸

5. Conclusion

To prevent market tipping, there is an urgent need to mandate sharing of user information in data-driven markets. Existing legal mechanisms to impose data sharing under EU competition law and data portability under the GDPR are not sufficient to tackle this problem. Mandated data sharing requires the design of a governance structure that combines elements of economically efficient centralization with legally necessary decentralization.

There are three feasible options. One is to centralize investigations and enforcement in a *European Data Sharing Agency*, while decision-making power lies with NCAs in a Board of Supervisors. The other, decentralized option is to set up a *Data Sharing Cooperation Network* to be coordinated through a *European Data Sharing Board*, with the NCA best placed to run the investigation serving as the *Lead NCA* that investigates and enforces the mandatory data-sharing decision across the EU. A third option is to mix both governance structures and to task national authorities with investigation (Lead-NCA) and decision making (DSCN) and the EU-level EDSA with enforcement of data sharing.

The advantage of this mixed option is that the technical infrastructure does not need to be duplicated across NCAs because it takes place at the EU level within EDSA. At the same time, no new investigation and enforcement powers have to be created at the EU level as NCAs select a Lead NCA, which is best placed to take on a particular case. The NCAs thereby jointly share the burden for the use of resources within the DSCN. Because of this combination of characteristics, the "Mixed" governance scheme seems to be most efficient.

Existing enforcement approaches have already shown the feasibility of such arrangements. By incorporating considerations regarding data protection and intellectual property protection into its very design, the governance structures proposed here provide a concrete approach towards the future regulation of data combining legal and economic insights that can be readily taken up by policymakers.

Author statement

Inge Graef has been mainly responsible for legal scholarship. Jens Prüfer has been mainly responsible for economic scholarship.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We thank Tobias Klein, Madina Kurmangaliyeva, Bertin Martens, Giorgio Monti, Patricia Prüfer and seminar participants at the Tilburg Law and Economics Center and at the Paris-Dauphine Working Group on Regulation of Platforms for helpful comments on an earlier draft. We have no conflicts of interest and acknowledge funding from the German Ministry of Finance (Grant No: fe 11/19). All errors are our own.

⁷⁸ Such a relationship of complementarity also exists with regard to the interaction between EU competition law and telecommunication regulation, as stated by the Court of Justice in Case C-280/08 *P Deutsche Telekom*, ECLI:EU:C:2010:603.

References

- Aldashev, G., Zananone, G., 2017. Endogenous enforcement institutions. *J. Dev. Econ.* 128, 49–64.
- Argenton, C., Prüfer, J., 2012. Search engine competition with network externalities. *J. Compet. Law Econ.* 73–105.
- Bernstein, L., Morrison, A., Ramseyer, J.M., 2015. Private orderings. *J. Legal Anal.* 7, 247–250.
- Bajari, P., Chernozhukov, V., Hortaçsu, A., Suzuki, J., 2019. The impact of big data on firm performance: an empirical investigation. *AEA Pap. Proc.* 109, 33–37.
- Bellovin, S.M., Dutta, P.K., Reiting, N., 2019. Privacy and synthetic datasets. *Stanf. Technol. Law Rev.* 22, 1–52.
- Bergemann, D., Bonatti, A., Gan, T., 2020. The Economics of Social Data. Yale University. Cowles Foundation discussion paper nr 2203 revised.
- Biglaiser, G., Calvano, E., Crémer, J., 2019. Incumbency advantage and its value. *J. Econ. Manag. Strategy* 28 (1), 41–48.
- Brynjolfsson, E., McAfee, A., 2012. *The Second Machine Age*. Norton and Company, New York.
- Calvano, E., Polo, M., 2021. Market power, competition and innovation in digital markets: a survey. *Inf. Econ. Policy* 54, 100853.
- J. Crémer, A.Y. De Montjoye and H. Schweitzer 2019. "Competition policy for the digital era," available at <https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf>.
- Drexel, J., 2017. Designing Competitive Markets for Industrial Data: Between Propertisation and Access. *JIPITECH*, pp. 257–292.
- Dixit, A.K., 2009. Governance institutions and economic activity. *Am. Econ. Rev.* 99 (1), 5–24.
- Edelman, B., 2015. Does Google leverage market power through tying and bundling? *J. Compet. Law Econ.* 11 (2), 365–400.
- European Commission. 2019. Press release "antitrust: commission opens investigation into insurance Ireland data pooling system," 14 May 2019.
- European Commission. 2020a "Communication 'a European strategy for data'," COM (2020) 66 final, 19 February 2020.
- Graef, I., 2016. *EU Competition Law, Data Protection and Online Platforms: Data as Essential Facility*. Kluwer Law International.
- Graef, I., Prüfer, J., 2018. Mandated data sharing is a necessity in specific sectors. *Econ. Stat. Ber.* 298–301.
- Graef, I., Husovec, M., Purtova, N., 2018. Data Portability and data control: lessons for an emerging concept in EU law. *Ger. Law J.* 1359–1398.
- A. Hagi, and J. Wright. 2020. "Data-enabled learning, network effects and competitive advantage," mimeo.
- Jin, G.Z., Wagman, L., 2021. Big data at the crossroads of antitrust and consumer protection. *Inf. Econ. Policy* 54, 100865.
- Jullien, B., Sand-Zantman, W., 2021. The economics of platforms: a theory guide for competition policy. *Inf. Econ. Policy* 54, 100880.
- Kathuria, V., Globocnik, J., 2020. Exclusionary conduct in data-driven markets: limitations of data sharing remedy. *J. Antitrust Enforc.* forthcoming.
- Klein, T., Kurmangaliyeva, M., Prüfer, J., Prüfer, P., 2021. A Simple Test for Data-Drivenness of Markets. Tilburg University mimeo.
- A. Lambrecht, and C. Tucker. 2015. Can Big Data protect a firm from competition? (December 18, 2015). Available at 10.2139/ssrn.2705530.
- Lundqvist, B., 2018. Competition and data pools. *J. Eur. Consum. Mark. Law* 146–154.
- Martens, B., 2021. An economic perspective on data and platform market power. In: *Proceedings of the JRC Digital Economy Working Paper 2020–09*.
- Masten, S.E., Prüfer, J., 2011. On the evolution of collective enforcement institutions: communities and courts. In: *Proceedings of the CentER Discussion Paper 2011–074*. Tilburg University.
- Masten, S.E., Prüfer, J., 2014. On the evolution of collective enforcement institutions: communities and courts. *J. Legal Stud.* 43, 359–400.
- Mayer-Schönberger, V., Cukier, K., 2013. The rise of big data. *Foreign Affairs*, May/June Issue.
- P. McAfee, J. Rao, A. Kannan, D. He, Q. Tao, and T.Y. Liu 2015. "Measuring scale economies in search," <http://www.learconference2015.com/wp-content/uploads/2014/11/McAfee-slides>.
- Prüfer, J., 2013. How to govern the cloud? *IEEE CloudCom* 33–38, 2013.
- Prüfer, J., 2018. Trusting privacy in the cloud. *Inf. Econ. Policy* 45, 52–67.
- Prüfer, J., Schottmüller, C., 2017. "Competing with Big Data", *TILEC Discussion Paper No. 2017-006*, CentER Discussion Paper No. 2017-007.
- Prüfer, J., Schottmüller, C., 2021. *Competing with big data*. *J. Ind. Econ.* forthcoming.
- M. Schaefer, G. Sapi, and S. Lorincz. 2018. "The effect of big data on recommendation quality: the example of internet search," *Düsseldorf Institute for Competition Economics Discussion Paper No 284*.
- Tombal, T., 2020. Economic dependence and data access. *Int. Rev. Intellect. Prop. Compet. Law*, 70–98.
- UK CMA. 2020. "Final report market study online platforms and digital advertising," 1 July 2020, available at https://assets.publishing.service.gov.uk/media/5efc57ed3a6f4023d242ed56/Final_report_1_July_2020.pdf.
- Varian, H., 2015. In: *Proceedings of the PowerPoint presentation at Lear Conference 2015 (June 25-26)*. <http://www.learconference2015.com/wp-content/uploads/2014/11/Varian-slides.pdf>.
- Veale, M., Binns, R., Edwards, L., 2019. Algorithms that remember: model inversion attacks and data protection law. *Phil. Trans. R. Soc. A* 376, 1–15.
- Williamson, O.E., 2005. The economics of governance. *Am. Econ. Rev.* 95, 1–18.
- Yakowitz, J., 2011. Tragedy of the data commons. *Harv. J. Law Technol.* 25 (1), 2–39.