Governance of Data Sharing

Graef, Inge; Prüfer, Jens

Publication date:
2021

Document Version
Early version, also known as pre-print

Discussion paper

GOVERNANCE OF DATA SHARING: A LAW & ECONOMICS PROPOSAL

by
Inge Graef and Jens Prüfer

22 January 2021

TILEC Discussion Paper No. 2021-001
CentER Discussion Paper No. 2021-004

ISSN 2213-9532
ISSN 2213-9419
https://ssrn.com/abstract=3774912

Tilburg University

Understanding Society

Governance of Data Sharing: a Law & Economics Proposal

Inge Graef and Jens Prüfer[1]

22 January 2021

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.*

Keywords: Data sharing, data-driven markets, economic governance, competition law, data protection, regulation

JEL Classification: K21, L44, L51, L86

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").[2]

[1] Graef: LTMS, TILEC & TILT, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands; I.Graef@tilburguniversity.edu. Prüfer: CentER, TILEC, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands; j.prufer@uvt.nl.

[2] Machine learning and artificial intelligence (AI) are not identical. Nevertheless, 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.

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*.[3] 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.[4] 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.[5] 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 competitors' user information.[6] 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.[7]

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 framework[8] 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 is considering to introduce a Data Act by the end of 2021 (European Commission 2020a). 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 has introduced an obligation for search engine providers qualifying as

[3] Argenton and Prüfer (2012) developed this definition of user information for the first time.
[4] 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.
[5] 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.
[6] See section 2 for details.
[7] 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., 2020).
[8] 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.

gatekeepers (namely 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 this obligation is too limited to address the concerns in this paper, its introduction does illustrate the relevance of data sharing in current regulatory discussions.

2. Why and When Mandate Data Sharing?

In a data-driven market, the interaction between a service provider and user is administered electronically such that it is possible to store the users' choices (e.g. clicking behavior) and characteristics (e.g. location) with very little effort. 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"). They show in a dynamic economic model 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.

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. 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

[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 capitalisation 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.

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 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.

Klein et al. (2020) follow up on the theory of Prüfer and Schottmüller. They 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. The next section develops a governance structure of data sharing that can be applied if this test indicates that a certain industry is data driven. 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)[11] 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.[12]

3. Governance of Data Sharing on Data-driven Markets

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

[11] This is the working paper version of Prüfer and Schottmüller (2021).
[12] Article 6(1)(j) of the proposed Digital Markets Act.

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.[13] 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) proposed 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. 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). 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.[14] 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 per cent market share*.[15] A market share threshold of 30 per cent 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 per cent and if the agreements do not contain certain types of severe restrictions of competition.[16] 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). Although our setting is different, this reasoning can be reversed such that a market share of 30 per cent indicates the ability of a

[13] Mandating only the sharing of user information is a lower bound on the amount of information shared. In personal conversation, one knowledgeable scholar noted: "I would put a research question mark next to that claim that this [mandatory sharing of algorithms] would hamper party's innovation incentives. I think there's little economic analysis about the optimal scope of trade secret protection—which is what this amounts to. We don't, for example, see the kinds of processes or optimization that we do (at least in principle) in patent and copyright." We leave it for future research to determine what else, besides user information, should be shared on a data-driven market.

[14] 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/).

[15] "Market share" in this sentence is to be understood broadly and flexible. Ideally, it refers to the share of users or another proxy for the amount of user information collected. 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.

[16] 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.

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 law[17] 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.[18]

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. 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.

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 Figure 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. 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 Figure 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

[17] Case C-62/86, *AKZO*, ECLI:EU:C:1991:286, par. 60.
[18] Article 6(1)(j) of the proposed Digital Markets Act.

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.[19]

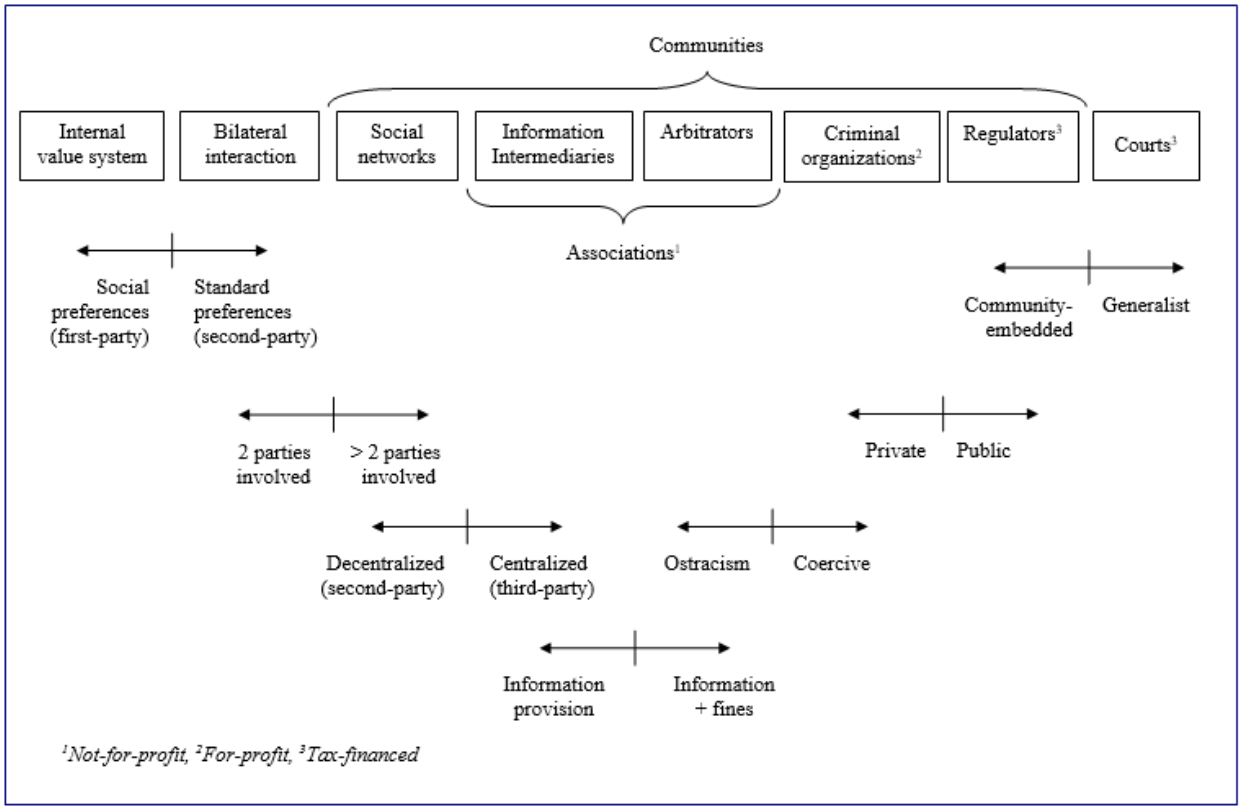

Figure 1: A Classification of Economic Governance Institutions (adapted from Masten and Prüfer, 2011).

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 a lengthy analysis short, however, 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

[19] 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.

monitor the operations of receivers, which is costly (and may furthermore infringe competition law).[20] 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 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,[21] 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) than a generalist judge who decides plainly based on the law. We will minimize this problem via a specific governance structure later.

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.**

What "regulators" can mean in the practice of data-driven industries and how to mitigate the moral hazard 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.

---

[20] 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

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

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. Moreover, our result of regulators as the best available institution implies that a *centralized* solution, where a third 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, the number of direct connections n*m can be very 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" 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*.[22]

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.[23] 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.

While it would be most efficient to create a new, independent EU body with fully centralized investigation and enforcement powers, the 1958 Meroni case law prohibits the delegation of discretionary powers to bodies not established by the EU Treaties.[24] A Treaty amendment would thus be necessary, which is a long and complicated process. The organization could instead be embedded within the European Commission's DG Competition. However, the competences of DG Competition are limited to the enforcement of the current EU competition rules.[25] As we discuss in section 4, the governance framework for data sharing goes beyond the remit of existing EU competition law.

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?

[22] The to-be-shared data would be collected from *n* firms, pooled centrally, and disseminated as a merged data set to *m* receivers.
[23] 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.
[24] C-9/56, *Meroni*, ECLI:EU:C:1958:7 and C-10/56, *Meroni*, ECLI:EU:C:1958:8.
[25] As laid down by Article 105 TFEU. Nevertheless, in its proposed Digital Markets Act the European Commission has granted itself the power to monitor new obligations for gatekeepers beyond the enforcement of EU competition law. In terms of data sharing, however, the scope of these obligations is too limited to address our concerns that go beyond the market for online search and the behaviour of large online platforms acting as gatekeepers. Furthermore, the institutional framework to enforce this new obligation in the Digital Markets Act is not as detailed as our proposal to implement data sharing in practice. See Article 6(1)(j) of the proposed Digital Markets Act.

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: first, the **creation of an agency at EU level** with autonomous investigation and enforcement powers under the control of the Member States and, second, a **network of national authorities** with coordination at the EU level.

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.[26] 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.[27] 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 (Figure 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.

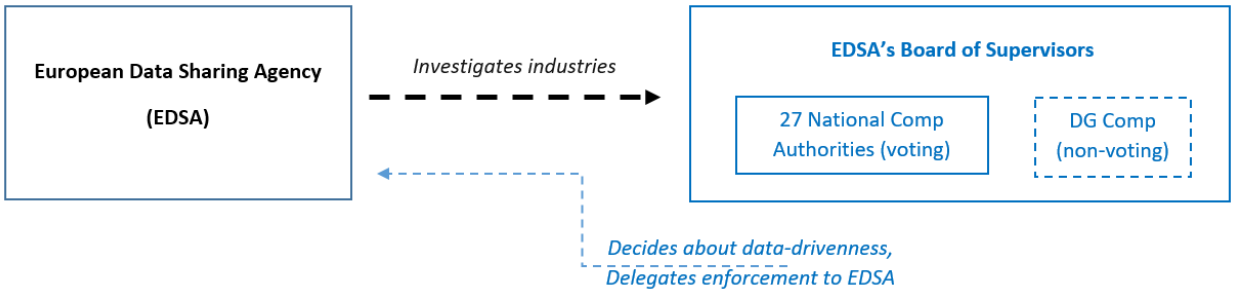

Figure 2: Governance of data sharing [Option 1]: The European Data Sharing Agency.

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.[28] 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 of the European Commission (non-voting).[29]

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

[26] 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.
[27] Case C-270/12, *UK v. Parliament and Council*, ECLI:EU:C:2014:18, par. 53-54.
[28] 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.
[29] Article 40 of Regulation (EU) No 1095/2010.

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.

Option 2: Establishing a Data Sharing Cooperation Network

Based on the enforcement of EU data protection and consumer law,[30] 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.[31] 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 (Figure 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 outs a recommendation containing the relevant markets that are in its view susceptible to ex ante regulation at the national level.[32] 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.[33] 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),[34] becomes

[30] 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.
[31] Compare Article 17(2) CPC Regulation and Article 56(1) GDPR.
[32] 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.
[33] Article 60(4) and 63 GDPR.
[34] Article 68(3) and (5) GDPR.

involved and can ultimately adopt a binding decision if the lead authority rejects objections raised by other authorities.[35]

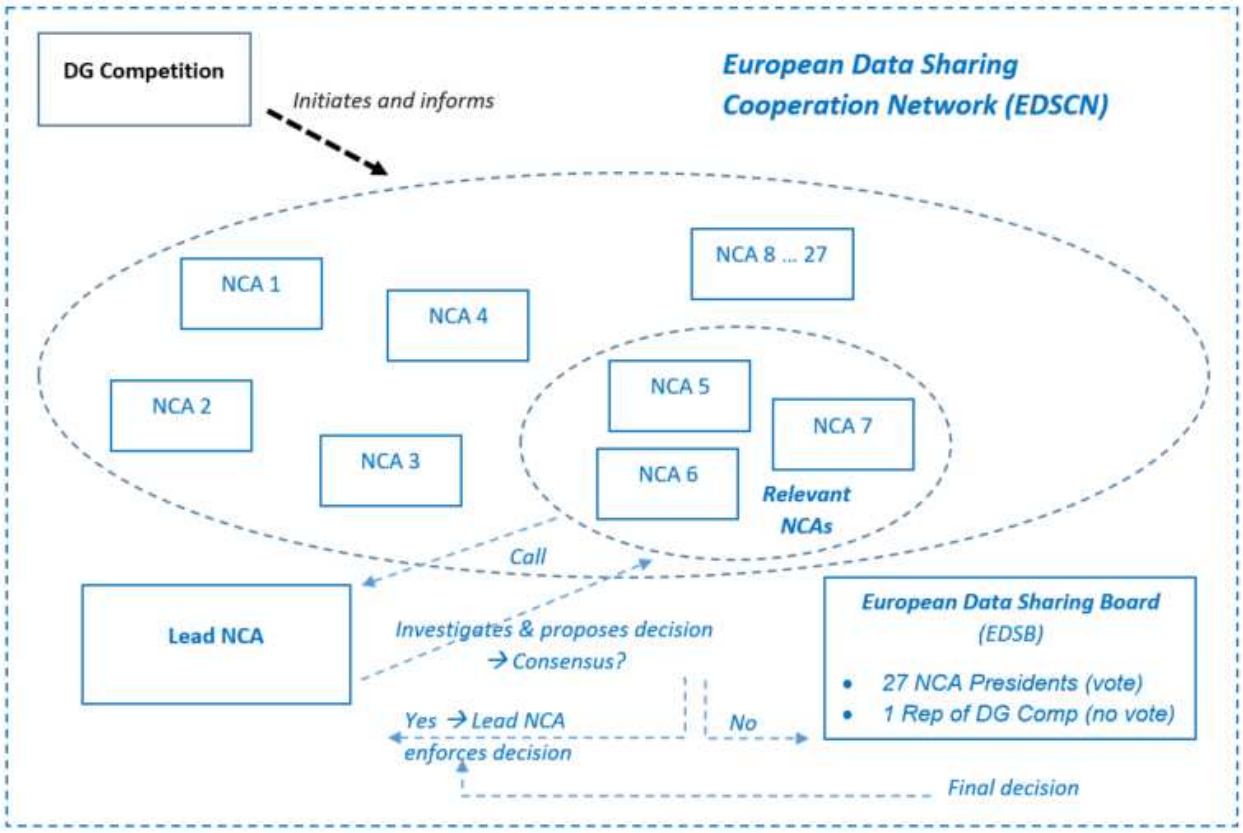

Figure 3: Governance of data sharing [Option 2]: The Data Sharing Cooperation Network.

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**.

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.

[35] Article 65(1)(a) GDPR.

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)

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

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 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.

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 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:[36]

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.[37]
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 possible[38] 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.[39]

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

---

[36] See for instance the discussion of anonymous use of individual-level data in Crémer, De Montjoye & Schweitzer (2019:85-87).

[37] 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. For more information, see https://en.wikipedia.org/wiki/Synthetic_data or, for an application, to https://www.syntho.ai/. 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/.

[38] 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.

[39] 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 (de 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).

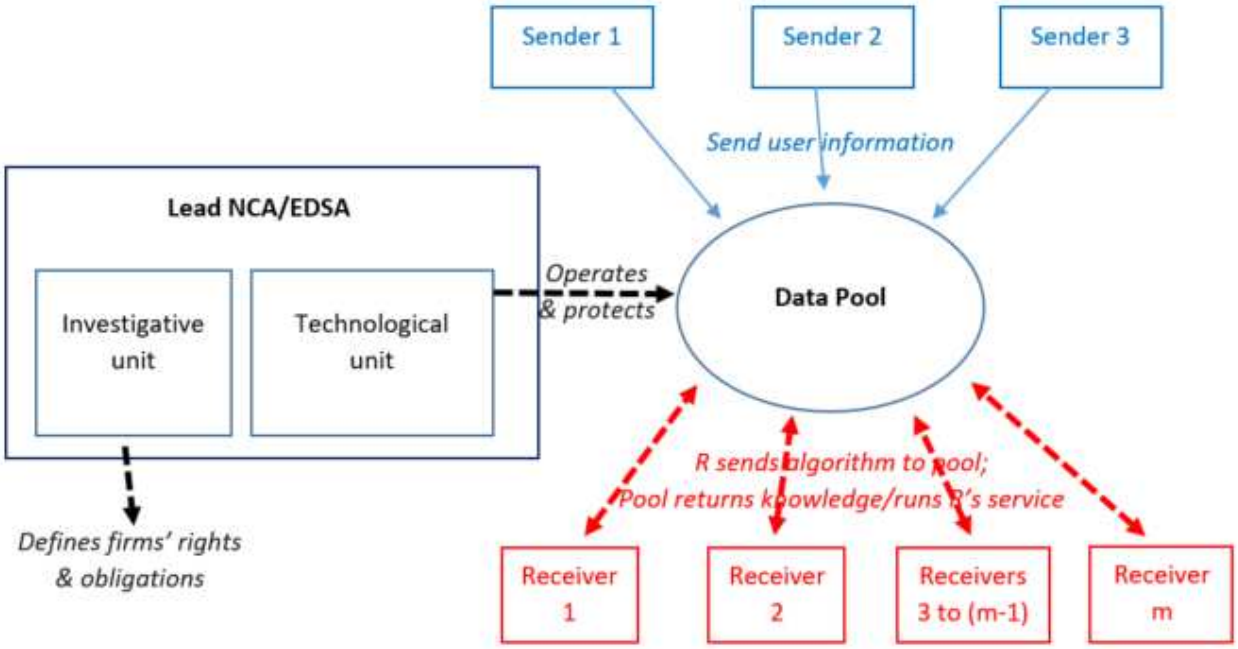

Figure 4: Implementation of Data Sharing on Data-driven Markets via a Data Pool.

The first advantage of this scheme is that data flows only via $n \leq 3$ links (instead of $n+m$ or even $n*m$) 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 eliminated. 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. Alternative Regulatory Options and Lessons 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.

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.

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 sharing, competitive risks can be constantly monitored.

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, and (3) an objective justification is absent.[40]

[40] 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. For refusals to license intellectual property rights the additional new product condition needs to be met, requiring a refusal to prevent the emergence of a new product for which potential consumer demand exists. In the Microsoft case, the General Court has accepted that a limitation of 'technical development' is sufficient. See Case T-201/04, Microsoft, ECLI:EU:T:2007:289, par. 647.

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 gas[41] and lottery services.[42] 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.[43] 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 & Schweitzer, 2019, p. 106). While previous cases met the high legal threshold under which a refusal to deal amounts to abuse of dominance,[44] 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.[45]

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)'.[46] 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 data subjects, which facilitates the exchange and re-use of personal data.[47] 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.

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.[48] Possible lawful grounds include consent of the data subject, performance of a contract,

[41] 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.
[42] 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*.
[43] 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.
[44] 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*.
[45] Article 6(1)(j) of the proposed Digital Markets Act.
[46] 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'.
[47] Article 20 GDPR.
[48] Article 5(1)(a) GDPR.

compliance with a legal obligation, and legitimate interests of the data controller.[49] 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 & 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 limitation[50] and data minimisation[51] 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.[52] 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 $n*m$ 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.

[49] Article 6(1) GDPR.
[50] Article 5(1)(b) GDPR
[51] Article 5(1)(c) GDPR
[52] Recital 50 GDPR. See the requirements to assess compatibility in Article 6(4) GDPR.

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.[53]

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 & 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, Husovec & Purtova 2018).

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* (Drexl, 2017). Copyright law protects the original expression of an idea and grants authors temporary exclusive rights.[54] While facts or data in itself do not qualify for copyright protection, databases 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.[55] In addition, data or information may benefit from trade secret protection if it is secret, has commercial value

[53] 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).

[54] 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.

[55] 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.

because it is secret and has been subject to reasonable steps to keep it secret.[56] 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.[57] In other words, trade secret protection cannot be invoked against the use of information obtained through legitimate means such as independent discovery or creation.[58]

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.[59]

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 minimisation and a balancing with the exclusivity offered by intellectual property rights).

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.

[56] 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.
[57] 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.
[58] Article 3 of the Trade Secrets Directive lists situations in which the acquisition, use and disclosure of a trade secret is considered lawful.
[59] 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.

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.

Acknowledgements

We thank Tobias Klein, Madina Kurmangaliyeva, Bertin Martens, Giorgio Monti, Patricia Prüfer and seminar participants at the Tilburg Law and Economics Center 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.

References

Aldashev, G. and Zanarone, G. 2017. "Endogenous Enforcement Institutions," Journal of Development Economics 128: 49-64.

Argenton, C. and Prüfer, J. 2012. "Search Engine Competition with Network Externalities," Journal of Competition Law and Economics: 73-105.

Bernstein, L., Morrison, A. and Ramseyer, J.M. 2015. "Private Orderings," 7 Journal of Legal Analysis: 247-250.

Crémer, J., De Montjoye, A.-Y. and Schweitzer, H. 2019. "Competition Policy for the Digital Era," available at https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf.

Drexl, J. 2017. "Designing Competitive Markets for Industrial Data: Between Propertisation and Access," JIPITECH: 257-292.

Dixit, A.K. 2009. "Governance institutions and economic activity," American Economic Review 99(1):5–24.

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.

European Commission. 2020b. "Communication 'Shaping Europe's Digital Future," COM(2020) 67 final, 19 February 2020.

Graef, I. 2016. "EU Competition Law, Data Protection and Online Platforms: Data as Essential Facility," Kluwer Law International.

Graef, I. and Prüfer, J. 2018. "Mandated data sharing is a necessity in specific sectors," Economisch Statistische Berichten: 298-301.

Graef, I., Husovec, M. and Purtova, N. 2018. "Data Portability and Data Control: Lessons for an Emerging Concept in EU Law," German Law Journal: 1359-1398.

Kathuria, V. and Globocnik, J. 2020. "Exclusionary conduct in data-driven markets: limitations of data sharing remedy," Journal of Antitrust Enforcement (forthcoming).

Klein, T., Kurmangaliyeva, M., Prüfer, J. and P. Prüfer. 2020. "A Simple Test for Data-Drivenness of Markets," mimeo, Tilburg University.

Lopez de Vallejo, I., Scerri, S., Tuikka, T. (eds). 2019. "Towards a European Data Sharing Space". Big Data Value Association: Brussels.

Lundqvist, B. 2018. "Competition and Data Pools," Journal of European Consumer and Market Law: 146-154.

Masten, S. E., and J. Prüfer. 2011. "On the Evolution of Collective Enforcement Institutions: Communities and Courts," CentER Discussion Paper 2011-074, Tilburg University.

Masten, S. E., and J. Prüfer. 2014. "On the Evolution of Collective Enforcement Institutions: Communities and Courts," Journal of Legal Studies 43: 359-400.

Prüfer, J. 2013. "How to Govern the Cloud?" IEEE CloudCom 2013: 33-38.

Prüfer, J. 2018. "Trusting Privacy in the Cloud," Information Economics and Policy 45: 52-67.

Prüfer, J., and C. Schottmüller. 2021. "Competing with Big Data," *Journal of Industrial Economics*, forthcoming.

Tombal, T. 2020. "Economic dependence and data access," International Review of Intellectual Property and Competition 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.

Williamson, O.E. 2005. "The Economics of Governance," 95 American Economic Review: 1–18.