

Jens Prüfer

# Competition Policy and Data Sharing on Data-driven Markets

Steps Towards Legal Implementation

**FOR A BETTER TOMORROW**

## FOR A BETTER TOMORROW

### A Friedrich-Ebert-Stiftung project 2018–2020

Growing social inequality, societal polarisation, migration and integration, the climate crisis, digitalisation and globalisation, the uncertain future of the European Union – Germany faces profound challenges. Social Democracy must provide convincing, progressive and forward-looking answers to these questions. With the project “For a Better Tomorrow”, the Friedrich-Ebert-Stiftung is working on recommendations and positions in six central policy areas:

- Democracy
- Europe
- Digitalisation
- Sustainability
- Gender Equality
- Integration

### Overall Coordination

**Dr. Andrä Gärber** is the head of the of Economic and Social Policy Division at the Friedrich-Ebert-Stiftung.

### Project Management

**Severin Schmidt** is a policy advisor for Social Policy in the Economic and Social Policy Division.

### Communication

**Johannes Damian** is an advisor of strategic communications for this project within the Department of Communications.

### The Author

**Jens Prüfer** is Associate Professor of Economics at Tilburg University and a member of the Tilburg Law and Economics Center (TILEC).

### Responsible for this publication at the Friedrich-Ebert-Stiftung

**Stefanie Moser** is a policy advisor for digitalization, trade unions and participation in the Economic and Social Policy Division.

**Dr. Robert Philipps** is a policy advisor for small and medium-sized enterprises and consumer policy in the Economic and Social Policy Division.



For further information please visit:  
[www.fes.de/fuer-ein-besseres-morgen](http://www.fes.de/fuer-ein-besseres-morgen)

# Competition Policy and Data Sharing on Data-driven Markets

## Steps Towards Legal Implementation

	<b>Foreword</b>	3
<b>1.</b>	<b>INTRODUCTION</b>	5
1.1	What is not proposed here .....	5
1.2	The problem: the natural tendency of data-driven markets towards monopolization .....	5
<b>2.</b>	<b>ALTERNATIVE PROPOSALS TO TACKLE TIPPING ON DATA-DRIVEN MARKETS</b>	8
2.1	Dominant firms should be broken up .....	8
2.2	Mandatory sharing of/open access to algorithms .....	8
2.3	Data portability .....	8
2.4	Tax data use .....	9
2.5	Competition law today .....	9
<b>3.</b>	<b>MANDATORY SHARING OF USER INFORMATION IN DATA-DRIVEN MARKETS: DETAILS OF A PROPOSAL</b>	10
3.1	How can a data-driven market be identified empirically? .....	10
3.2	What information exactly should be shared on which market? .....	11
3.3	How to anonymize user information and how to avoid re-identification of individuals .....	12
3.4	Who exactly should share data? .....	12
3.5	Who should have the right to obtain access to the shared data? At what price? .....	14
3.6	How should data sharing be organized? What is the optimal govern- ance structure? .....	14
<b>4.</b>	<b>CONCLUSION</b>	16
	Table of figures	18
	References	19



## Foreword

Over 90 per cent of internet searches go through Google; in social media Facebook's European market share is over 70 per cent; and almost half of Germany's online commerce now takes place via Amazon.<sup>1</sup> A few companies dominate broad swathes of internet commerce. This tendency towards monopolisation not only endangers competition, but in the medium term weakens business's innovative capacity. Studies also show that market concentration goes hand in hand with inequality or income and wealth (Ferschli et al. 2019; Autor et al. 2019).

European policymakers are largely at one on the need for regulatory action. They remain divided on how the problem can best be addressed, however. Most proposals rely on traditional («legacy») competition-law instruments or even the break-up of the tech giants. But can the regulatory challenges of the digital age be solved solely with existing cartel law or do we need new, innovative policy solutions?

A still fairly recent proposal that is increasingly attracting attention in Europe is the data sharing obligation. The SPD first floated this idea in February 2019 within the framework of an initiative for a «data-for-all law» (SPD 2019a) and substantiated the proposal in a resolution at the party congress later that year (SPD 2019b). There is similar thinking at EU level (Government of the Netherlands 2019). The data sharing obligation is based on the assumption that data or access to data is key to solving the problem. Companies that dominate a data-driven market should be obliged to share their data – with other firms active in the same market, but also with public and civil society organisations. According to advocates, this would give rise to fairer competition on the digital market, which in turn will spur innovation. The data sharing obligation could also breathe life into the idea of data as a «common good», which is both produced and used by society.

In a number of respects, the data sharing obligation takes policymakers into new legislative territory. Numerous questions arise, accordingly. Under what conditions will a company have to share its data? Who will have access to this data? Which data will have to be shared and how can it be shared without infringing other legal provisions, in particular data

protection? And last but not least, what technical and institutional infrastructure are required for data sharing?

To clarify these issues we asked Jens Prüfer, Associate Professor for Economics at the University of Tillburg and one of the prime movers of the data sharing obligation, to develop some recommendations on how the idea can be implemented in legislative practice. Also incorporated in this process are the results of an expert discussion held in summer 2019 within the framework of which we discussed implementation options with representatives from business, politics and civil society. We would like to thank all participants for their contributions.

We hope that the present study can help to clarify the key questions pertaining to the data sharing obligation and make it more tangible for policymakers and business.

---

<sup>1</sup> For statistics on Google and Facebook for December 2019, see Statcounter (no date a) and Statcounter (no date b); for Amazon's market share for 2017 see University of St. Gallen (no date).



## 1

# INTRODUCTION<sup>2</sup>

We are drowning in data. Around 90 per cent of the world's data today was created in the past two years.<sup>3</sup> Most of it is unstructured text, images and videos, which is hard to categorize, let alone understand, for human beings.<sup>4</sup> There are sensor data in (self-driving) cars, smart home and office equipment, social media data, mobile data, data on internet and browsing behaviour, or digital camera images, to name just a few. This explosion of data is accompanied by tremendous progress in data science methods, which can make sense of all the available information. These methods are fuelled by artificial intelligence (AI). The McKinsey Global Institute recently projected that the adoption of AI by firms may follow an S-curve pattern: a slow start given the investment associated with learning and deploying the technology, and then acceleration driven by competition and improvements in complementary capabilities (Bughin et al., 2018). At the macro level, they expect that AI could potentially deliver additional economic output of around U\$13 trillion by 2030, boosting global GDP by about 1.2 per cent a year.

Realizing these potential gains, however, requires a couple of decisions we need to take. At their core is the insight that the surge of big data and AI we are witnessing is not exogenous: it is the consequence of human actions that are driving technological progress. And these actions, taken mainly by businesses in the United States, are the outcome of an economic and legal system that enables and incentivizes entrepreneurs to innovate. Hence, what we need to ask, if we want to understand how to let as many people as possible enjoy the benefits of technological progress, is the following:

How can we shape innovation-support institutions such that the incentives to invest in innovation are high and ensure that the benefits that stem from mere by-products of innovation are distributed widely, for many to enjoy?

Such successful institutional design would avoid extreme inequality of opportunities and outcomes and, hence, contribute to social cohesion and welfare in these times of great technological disruption.

## 1.1 WHAT IS NOT PROPOSED HERE

It has been speculated that the mandatory »sharing of data« could solve many problems of our »datafied« society.<sup>5</sup> However, such general claims leave many questions unanswered and lack a theory of harm. At this level of generality, the idea is terrifying to many companies, who fear that their business secrets would be exposed to competitors with better-developed analytic capabilities (or that those secrets could be inferred from mandatorily shared datasets related to their technologies and operations). It is also threatening to many consumer and privacy protectors, who fear that citizens' personal data, together with individual identifiers, would be publicly exposed and hence enable large-scale privacy intrusion, from both legal and illegal actors.

While a general, unrestricted data-sharing obligation may have tremendous negative consequences, which are not the subject of this essay, I will explain when and why a certain kind of data sharing would have positive consequences in specific sectors. Summarizing those positive consequences, given today's knowledge, the specific kind of data sharing discussed below seems the best solution available to tackle a genuine problem on data-driven markets: monopolization with the consequence of diminished incentives to innovate both for dominant firms and their (potential) competitors. If it is not implemented (the sooner the better), the total benefits of datafication (= big data + AI) described above will be reduced and enjoyed ever more unequally, with all the dismal social, political and economic effects of excessive inequality.<sup>6</sup>

<sup>2</sup> Jens Prüfer: I am grateful to my colleagues at the Tilburg Law and Economics Center (TILEC) for valuable feedback, especially Francisco Costa-Cabral, Inge Graef, Tobias Klein, Madina Kurmangaliyeva, Giorgio Monti and Patricia Prüfer. Constructive comments by Stefanie Moser, Robert Philipps and Bastian Jantz are also appreciated. All errors are my own.

<sup>3</sup> <https://public.dhe.ibm.com/common/ssi/ecm/wr/en/wr12345usen/watson-customer-engagement-watsonmarketing-wr-other-papers-and-reports-wr12345usen-20170719.pdf>

<sup>4</sup> <https://blog.microfocus.com/how-much-data-is-created-on-the-internet-each-day/>

<sup>5</sup> SPD (2019), Digitaler Fortschritt durch ein Daten-für-Alle-Gesetz, [https://www.spd.de/fileadmin/Dokumente/Sonstiges/Daten\\_fuer\\_Alle.pdf](https://www.spd.de/fileadmin/Dokumente/Sonstiges/Daten_fuer_Alle.pdf)

<sup>6</sup> Mayer-Schönberger and Ramge (2018) provide data on and analysis of this inequality.

## 1.2 THE PROBLEM: THE NATURAL TENDENCY OF DATA-DRIVEN MARKETS TOWARDS MONOPOLIZATION

In the past two decades, we have witnessed the unprecedented, indeed stellar rise of a few companies that offer products and services that billions of users want to use and which have changed our lives in many respects. They have been rewarded with equally unprecedented profits and stock-market valuations, sometimes – temporarily – exceeding US\$1 trillion for the top firms, reflecting shareholders' expectations about their future profitability. Such expectations are not unfounded.

These »superstar firms« (Autor et al. 2017) have been remarkably successful in identifying users' needs, in developing devices and services that satisfy demand – and in understanding and exploiting fundamental economic mechanisms that make these markets so special. What distinguishes the superstar firms from others is that they understood early,<sup>7</sup> first, that data are the key ingredient boosting the value of their services for users and second, accompanying this insight, the economics of data-driven markets.

In a data-driven market, the interaction between a service provider and a user is administered electronically such that it is possible to store users' choices (for example, clicking behaviour) and characteristics (for example, IP address, language preferences, or location) with very little effort. Examples include search engines, digital maps, platform markets (for example, for hotels, transportation, dating, music/video-on-demand); probably also smart meters, self-driving vehicles and various other markets. Due to progress in computer science and in data analytics, analysing »big data« and making predictions about individuals' or groups' future behaviour has become much more effective in the past few years.<sup>8</sup>

Prüfer and Schottmüller (2017) call the information about users' preferences and/or characteristics user information.<sup>9</sup> They show in a dynamic economic model that, in data-driven markets, user information is a key input in the innovation process: via feedback effects (»data-driven indirect network effects«), user information leads to market tipping (monopolization; see

<sup>7</sup> In fact, (only) since 2001 has Google made use of the data created by logging users' clicking behaviour in so-called »search logs« (Zuboff 2016). Before that, Google's original page rank algorithm, as described in Page et al. (1999), let the rank of a website be determined only by the structure of hyperlinks on the world wide web, not by user information.

<sup>8</sup> See, for example, Le Cun et al. (2015). A very accessible introduction to the economic consequences of better prediction is Agrawal et al. (2018).

<sup>9</sup> To be clear: user information refers to all information that a service provider gains about an individual user or a group of users by logging interactions with that user. It excludes data about the quality of machines (for example, the workings of certain technologies within a car) but includes data about the people in the car (for example, the reaction speed of the driver or the choices of people on the backseats regarding videos shown by the in-car entertainment system). Data generated by CCTV cameras about the number and walking speed of pedestrians or the identity of those pedestrians, generated by an accompanying face recognition system, are also regarded as user information; but information about how long the CCTV camera can run after an energy fallout is not. Hence, the concept of user information is much broader than the data sometimes requested by websites for registration, for example, name, (e-mail) address or phone number.

Box 1). 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.

This tendency of data-driven markets to tip 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 that the dominant firm has access to 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 once again in the runners-up spot, which entails few users and low revenues, but it would still have to pay the large costs involved in attempting a leap in innovation. Foreseeing this situation, no rational entrepreneur would invest in innovation in a smaller firm. In turn, because the dominant firm knows about the disincentive to innovate among its would-be competitors, it is protected by its large (and constantly renewed) stream of user information and can remain content with a lower level of innovation, too.

### Box 1

#### MARKET TIPPING

In principle, every market can be dominated by a single firm. In economic theory, we usually speak of a monopoly even if the »monopolist« does not have 100 percent market share but is only very dominant and the other firms in the market cannot affect the price that is accepted by consumers very much. If a market is moving towards monopoly, we speak of market tipping: the market share of the dominant firm is continuously increasing, while other firms' market shares are dropping. Market tipping can have various reasons, including static or dynamic economies of scale or various network effects: direct network effects (a user's utility increases in proportion to the number of other users, for example, in telecommunications) and indirect network effects (a user's utility increases in proportion to the number of users on the other side of a market, for example, dating platforms) have been studied. Data-driven indirect network effects, however, are a novel phenomenon driven by big data (for example, for search engines): a user's utility is not directly affected by other users but the amount of other users increases the provider's stock of user information, which in turn decreases the cost of innovation for the provider. All else being equal, this provides the dominant company with an initial lead in innovation, which – again – increases its access to user information and so grows into a self-reinforcing competitive edge (and finally market tipping). Once a market is tipped, however, the incentive to innovate decreases sharply for all market players (including the dominant firm).



Such relatively low innovation rates, both by the dominant firm and by (would-be) competitors, compared with a situation with lively competition, constitute the theory of harm in Prüfer and Schottmüller (2017). Moreover, they 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 are related to geographic information. These data are also valuable when providing a customized map service. The authors showed that if the market entry costs 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 of this process, which can occur repeatedly, generating a domino effect.

The third contribution of that paper was, based on the earlier idea of Argenton and Prüfer (2012), to study the consequences of a regulatory requirement for dominant firms in data-driven markets to share their (anonymized) data on user preferences and characteristics with each other. Prüfer and Schottmüller 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 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 in that market.

Now, before addressing critical issues surrounding the practical implementation of the mandatory data-sharing proposal in Section 3, I will first sketch why several alternative proposals mentioned in previous discussions are not well suited to tackle the tipping problem on data-driven markets.

## 2

# ALTERNATIVE PROPOSALS TO TACKLE TIPPING ON DATA-DRIVEN MARKETS

Because the market power of dominant firms on various data-intensive markets has grown over the past few years, regulators, competition authorities and politicians in many countries have started to question the roots and consequences of this development (see Section 4). Several proposals have been made on how to tackle the strong and persistent dominance of big tech firms in such markets. Here, I will briefly discuss a few of them.

### 2.1 DOMINANT FIRMS SHOULD BE BROKEN UP<sup>10</sup>

The specific market-tipping problem on data-driven markets stems from the often inseparable connection because offering a service/selling a good and logging the user's choices and characteristics during their consumption/purchase. Consequently, if a dominant firm on such a market was forced to divest parts of its business, there would still be one part left that continues to run the core service, to interact with users, and thereby to amass exclusive user information – and hence to tip the market in the long run. The same would even be true if the firm was mandated to split its core service, for example, running a search engine, into two identical and competing parts: if the sharing of algorithms, user information, and other resources occurred only once, only one part could occupy the same web spot (URL) as the previous monopolist, which would give it a slight head start. Then, due to the nature of machine-learning (self-adapting) algorithms and the fact that users can interact only with one provider for a given service/purchase at a given point of time (for example, to get one search query answered), the part with the head start would get more user information than the other. Instantaneously, the market-tipping dynamic, as analysed in Prüfer and Schottmüller (2017), would start again. The problem would not be solved.

### 2.2 MANDATORY SHARING OF/OPEN ACCESS TO ALGORITHMS

If a complete breakup of dominant firms is unlikely to happen, one may be tempted to require that their matching or predic-

tion algorithms be opened up so that competitors can take a look and learn from the dominant firm. This may be an effective strategy in a static economy without a future, but it is doomed to fail in a dynamic high-tech industry that competes by innovation. The reason is that innovating by improving algorithms is costly at the margin; that is, every additional »unit of innovation effort« – for example, hiring another researcher – has to be paid for. Consequently, if a firm (the dominant one or another one) knows that the fruits of its innovation investments today may be taken away tomorrow because it may be forced to share its insights concerning its algorithms with its competitors, its propensity to spend on innovation today will decrease steeply.<sup>11</sup> Because this effect counteracts the goal of increasing innovation levels in data-driven markets, the forced opening of algorithms would be an inappropriate policy measure.

### 2.3 DATA PORTABILITY

Article 20 of the EU's General Data Protection Regulation (GDPR) provides users of a service with the right to ask their provider for their personal data and to transfer it to another (competing) provider (data portability). Here I will not discuss the pros and cons of this legal provision (see Graef et al., 2018, for a competent treatment), but only point to the fact that data portability cannot solve the problem of market tipping on data-driven markets. The reasons are threefold.

First, even if a user requests their personal data from a provider, the provider is not obliged to delete the data and, hence, a dominant firm does not lose data even if all its users were to make use of their right to data portability. Second, the GDPR does not extend to insights that are derived, often by using AI, from combining the personal data sets of millions of users. If, via data portability requests, a competitor gets hold of only a fraction of individual sets of personal data, compared with the dominant firm, the dominant firm would still be able to learn more, for example, about which product features are especially sought after by which type of users. This would make the value of one additional dataset for a competing firm smaller than for the dominant firm. Finally and most importantly, users' data portability requests are made

<sup>10</sup> See, for example, <https://edition.cnn.com/2019/03/08/politics/elizabeth-warren-amazon-googlefacebook/index.html>

<sup>11</sup> This is the core problem of every (expected) expropriation. If property rights are insecure, incentives to invest in the improvement or maintenance of resources are shallow.

in parallel and for each individual separately. Invoking one's data portability right comes at positive marginal cost because each individual user has to contact the provider and request the data, verify their identity and then transfer it to a competitor. Given that the value of the personal information of one more user is extremely small (if one already has millions of users), a user cannot expect that the transfer of their data will increase the quality of their new provider in a recognizable way. Consequently, it is incentive-compatible for many users not to invest the effort into requesting their data. Therefore, it can be expected that data portability will remain rather toothless to mitigate the tipping of data-driven markets.

## 2.4 TAX DATA USE

Due to the very asymmetric distribution of the gains generated by dominant firms on several platform markets, many of which are driven by data, suggestions have been made to tax the use of data. While there may be good reasons for such a policy from a distributional-justice perspective, such a policy would not tackle the main problem on data-driven markets, market tipping. The reason is that, on one hand, generating and using data can be efficient as it allows service providers to match users better with information or services they like. Reducing this benefit by increasing the price of generating or transferring data, via taxation, would decrease the incentives to collect and use data and, hence, decrease service quality as perceived by users/consumers. On the other hand, because market tipping on data-driven markets is a consequence of competitors' asymmetric access to key resources, most notably user information, and as taxation would not affect the relative lead of the dominant firm in accessing user information, this policy would be unable to tackle market tipping.

## 2.5 COMPETITION LAW TODAY

In many discussions about the law and economics of data-driven markets, scholars and others assert that today's competition law, most notably Article 102 TFEU, which prohibits the abuse of a dominant position in a market, is sufficient to avoid major problems. In my opinion, existing EU competition law falls short of the goal of increasing competition and innovation.

Article 102 TFEU is a powerful tool if a dominant firm actually engaged in conduct that abuses its market power and hence leverages it to other products or forecloses markets. It can also help to punish uncompetitive behaviour in data-driven markets.<sup>12</sup> But competition law works *ex post* and case-based: every time a competitor asks a dominant firm to share its user information, based on the claim that those data are an essential facility to compete in this market, a new case would be established and have to be decided by the relevant competition authority and, potentially, appeals courts. As

shown by Prüfer and Schottmüller (2017), however, market tipping occurs in data-driven markets even if the dominant firm's conduct is flawless as it depends on data-driven indirect network effects, which are an unavoidable (and potentially very efficient) economic characteristic of such markets. Moreover, if a market is found to be data-driven, tipping can be predicted. Abusive behaviour by a dominant firm is not necessary to tip the market or to discourage competitors from innovating heavily. Therefore, what is needed is the option to intervene in such markets *ex ante*, that is, before the market has tipped and the dominant firm can be accused of abusing its position (which perhaps it has not). Moreover, in order to avoid the cumbersome and lengthy process of repetitive legal cases, a quicker and more flexible tool that allows competition authorities to intervene in markets without invoking the essential facilities doctrine is needed, a tool that is akin to regulation once a market has been identified as data-driven.<sup>13</sup>

<sup>12</sup> See the Google Shopping case for a good example: [http://ec.europa.eu/competition/elojade/isef/case\\_details.cfm?proc\\_code=1\\_39740](http://ec.europa.eu/competition/elojade/isef/case_details.cfm?proc_code=1_39740)

<sup>13</sup> The Dutch Ministry of the Economy recently published an open letter proposing adjustments of competition law. They include the option for mandatory data sharing in specific industries, *ex ante* interventions, and several other interesting ideas. In English: <https://www.government.nl/latest/news/2019/05/27/dutch-governmentchange-competition-policy-and-merger-thresholds-for-better-digital-economy>. In German: <https://www.government.nl/documents/letters/2019/05/23/zukunftstauglichkeit-des-wettbewerbsinstrumentariums-in-bezug-auf-onlineplattformen>

# 3

## MANDATORY SHARING OF USER INFORMATION IN DATA-DRIVEN MARKETS: DETAILS OF A PROPOSAL

The contributions of Prüfer and Schottmüller (2017) are theoretical. They suggest explanations of the empirical patterns we observe, but they are short on giving detailed hints about how their ideas, especially the data-sharing policy proposal, could be translated into practice. They declare that: On a data-driven market, due to indirect network effects driven by machine-generated data about user preferences or characteristics, marginal costs of innovating are decreasing in demand. The policy proposal is adapted from Argenton and Prüfer (2012) and stated as follows: On a data-driven market, all user information should be anonymized and shared by all competitors with each other.

In practice – and hence for the purpose of this study – the following questions need to be answered:

- 1 How to identify a data-driven market empirically?
- 2 Precisely what information should be shared on which market?
- 3 How can user information be anonymized and how can re-identification of individuals (technically or legally) be avoided?
- 4 Who exactly should share data?
- 5 Who should have the right to get access to the shared data? At what price?
- 6 How should data sharing be organized? What is the optimal governance structure?

I will address these questions in order of appearance.

### 3.1 HOW CAN A DATA-DRIVEN MARKET BE IDENTIFIED EMPIRICALLY?

One of the most challenging tasks when designing a mandatory data-sharing law is the delineation of its domain of applicability. To start with, there must be some kind of test that determines whether an industry is data-driven or not. This is important because Prüfer and Schottmüller (2017) showed that the net welfare effects of mandatory data sharing are unambiguously positive only if data-driven indirect network effects are sufficiently pronounced, that is, if user information decreases the marginal cost of innovation substantially and not only modestly. In the latter case, data sharing can still be positive but, due to the resulting multiplication of innovation costs when several providers compete actively with each other, the net welfare effect is unclear.

Such a test ideally produces an index of data-drivenness, It is necessarily applied at the industry level. Its result would be that, for instance, »industry A has a high degree of data-drivenness and therefore mandatory data sharing is warranted«, whereas »industry B is only mildly data-driven such that there should be no mandatory data sharing«.

Any exercise at industry level requires some kind of market definition. As we know from merger regulations, for instance, any market definition (or delineation) is cumbersome and easy to criticize that increases the transaction costs of the entire exercise. Moreover, some people argue that many data-driven markets overlap (for example, that Google, Facebook, Microsoft and Apple are competing in the online advertisement market). Therefore, a reasonable starting point for market definition is to take a user/consumer perspective and to ask what service that user demands now and which providers may have an offer that can satisfy the user's needs. For instance, users usually have no specific demand for advertisements, but they can well specify whether they are looking for the best match with a general search query (hence the relevant market would be general search engines) or a hotel in Berlin (relevant market: overnight stays at hotels in Berlin) or the route to that hotel (relevant market: route planners with information on the way to Berlin). In none of these cases is the relevant market »advertisement«.

As for the test for data-drivenness, it is important to understand the novel quality of data-driven indirect network effects, compared with other economic effects. They combine the value of more user information on the demand side of a market with lower marginal cost of innovation on the supply side. Therefore, a test for data-drivenness of a market should include (at least) two parts. First, to understand the demand side better, we should empirically analyse – for example, by administering randomized field experiments within or between firms from the respective industry – the extent to which the artificial interruption of access to user information decreases service quality, as perceived by users. One could administer user surveys to learn the extent to which the artificial quality decrease during an experiment was a mere nuisance (and dominated by other product characteristics, such as service speed or user friendliness) or when it annoyed users so much that they reconsidered their choice of service provider.

Second, to understand the supply side, one could develop quantitative measures of product quality, such as the speed of delivering a service.<sup>14</sup>

In data-driven markets, the analysis would reveal an endogeneity issue: access to more user information increases quality measures; higher product/service quality increases demand for the service, which, in turn, increases the amount of user information even more.<sup>15, 16</sup> If such a circle can (not) be empirically identified, the market is (not) data-driven.

In practice, such tests will be neither easy nor quick. Therefore, it is important that, in the long run, the organization that is in charge of enforcing the mandatory data-sharing law (for example, a new EU-level agency or a new division of DG Comp; see Section 3.6) has enough legal, human, and financial resources to administer such tests at the level of many industries, starting with the prime suspects.

Complementing this procedure, the idea of turning around the burden of proof could be applied, as was promoted in the recent Vestager Report for the EU's DG Competition (Cremer et al., 2019). Even if the original proposal is related to cases concerning the abuse of a dominant position, the underlying logic is straightforward and can inform our current question. If, after some pre-testing (see suggestions above), a firm is deemed to be dominant in a potentially data-driven market and the firm objects to this claim, it would be up to that firm to show that it is not dominant. Given that firm's superior access to user information and other characteristics of its product, this obligation seems more efficient than the obligation on an outsider – for example, a competition authority – to prove dominance without access to such information. Box 2 summarizes these considerations.

Box 2

### POTENTIAL SEQUENCE OF ACTIONS TO ESTABLISH THE DATA-DRIVENNESS OF A MARKET

- 1 market definition (user centric);
- 2 study the demand side of the market: what drives users' consumption utility?
- 3 study the supply side of the market: what drives »objective« measures of product quality?

<sup>14</sup> See He et al. (2015) or Schaefer et al. (2018) for attempts to measure the quality of general search engines and its dependence on access to user information (in the form of search log data).

<sup>15</sup> In »traditional« markets that are not data-driven, higher quality also leads to higher demand, *ceteris paribus*, but higher demand does not boost product/service quality, at least not to the extent that investments in other aspects of R&D, such as hiring more researchers, could be substituted by the gains that come through access to more demand (and user information).

<sup>16</sup> This concept is related (but not identical) to the test for small but significant non-transitory decrease in quality (SSNDQ), which is modelled after the better-known SSNIP test. See Gebicka and Heinemann (2014:158) for details.

## 3.2 WHAT INFORMATION EXACTLY SHOULD BE SHARED ON WHICH MARKET?

The goal of the policy proposal at hand is to avoid market tipping on data-driven markets or, in case a market has already tipped, to make it contestable again. For that purpose, only the sharing of user information is relevant, no other data.<sup>17</sup> These are raw data about users' choices or characteristics, which can be automatically (and hence at virtually zero marginal cost) logged during a user's interaction with a service provider.<sup>18</sup> The policy proposal explicitly does not include processed data, in which the original service provider/dominant firm already invested effort, for instance for data analytics (and hence at positive marginal cost). If such data would be required to be shared, it might facilitate free-riding of smaller competitors and crowd out the dominant firm's incentives to invest into analytics in the first place (see footnote 12). If only raw data are shared, it also incentivizes competitors to develop own analytics techniques, which can lead to a plurality of approaches, differentiated products, and, hence, more choice for consumers.

It may be tempting to require also the sharing of other data, but such requirements would first warrant a separate theory of harm for the case in which they are not shared. In contrast, the proposal at hand is strictly designed to mitigate market tipping on data-driven markets, which appears to be the main problem on such markets. All other issues discussed in Section 2, including the potential to abuse a dominant position or the possibility to earn huge monopoly profits and have them taxed elsewhere, decrease once a market is more competitive.

Finally, in order to disseminate the effects of data sharing as quickly as possible, user information would have to be shared on a »continuous« basis, that is, as frequently as technically possible.<sup>19</sup>

<sup>17</sup> Specifically, there is no need to ask anybody to share technical data about machines or business processes that go beyond the immediate logging of users' choices and characteristics, as known to the respective service provider.

<sup>18</sup> Like all types of information, user information is non-rival, but excludable. By lowering the hurdles of excludability via a mandatory data-sharing law, its attributes become closer to a public good, the sharing of which is efficient. For instance, in the search engine industry, this user information would be contained in search log files. Schaefer et al. (2018:5) describe the data they retrieved from Yahoo! as »information about the search term, the time when the search term was submitted, the computer from which the search term originates, the returned list of results, and the clicking behavior of the user«.

<sup>19</sup> Recently, Google, Facebook, Microsoft, Twitter and other firms showed that sharing of user data is technically and organizationally possible at a large scale and automatically. 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-microsofttwitter>.

### 3.3 HOW TO ANONYMIZE USER INFORMATION AND HOW TO AVOID RE-IDENTIFICATION OF INDIVIDUALS

Critics object to the potential trade-off between more competitiveness in the market and less individual privacy if data are shared. There are legitimate concerns regarding the data protection implications of mandatory data-sharing, which have to be addressed. User information that can be linked to a specific user qualify as personal data under the GDPR, which dictates that such information cannot be shared without a legitimate ground for data processing, such as previous consent of users.

Fortunately, privacy-related reservations about data sharing can be mitigated by the following considerations:

Data sharing is unproblematic under the GDPR regulation as long as the data shared are anonymized. Anonymization irreversibly destroys any way of identifying the data subject, therefore anonymized data cannot be re-linked to an individual user.

But even when user information is anonymized, there is a risk of »indirect re-identification«, when shared data are matched with other data sources. However, technological options exist to make the re-identification of a specific individual expensive and hence less attractive. Furthermore, in order to mitigate the risk of indirect re-identification a data-sharing law could be introduced which prohibits the re-identification of individuals from shared data. Thus re-identification would establish an illegal act and would be punished.

Under a data sharing regime that effectively prohibits the re-identification of data subjects, pseudonymizing data might also be considered (see Box 3 below). However, to which extent this is a viable option has to be carefully examined together with data protection authorities and other stakeholders.

Privacy and data security concerns could be further mitigated by building data protection into the data sharing governance structure (see Section 3.6). Thus, instead of granting organizations direct access to shared data, a trustworthy data intermediary (data trustee) could be established to be responsible for guaranteeing the compliance of data sharing with data protection rules. The intermediary – for instance, a data protection authority – could safely pool shared user information and ensure the fundamental privacy rights of the users by anonymizing the data before sharing it with eligible third parties.

This idea could be taken even further: if there was a data trustee, it would be possible to pool all shared user information, independent of whether it is personal or non-personal data, »behind a curtain«, that is, on a server where no human being has access to the data. Instead of sharing anonymized data with the organizations eligible under the data sharing regulation, those organizations could unleash their (differentiated, competing) algorithms to use that data pool as training data, without receiving any of the shared data themselves. The results, meaning the trained algorithms, but not the originally shared user information, would be transferred back to the providers who could offer their servic-

es to users on a competitive basis. In May 2019, Finland passed legislation on the secondary use of health and social data that envisions such an infrastructure and process that already comes close to the idea outlined above.

Box 3

#### ANONYMIZATION VS. PSEUDONYMIZATION

Anonymization refers to the process of either encrypting or removing personally identifiable information from datasets, such that the people whom the data describe (data subjects) remain anonymous.

Pseudonymization refers to the process of replacing personally identifiable information fields by one or more artificial identifiers, or pseudonyms. Notably, pseudonymized data can be restored to its original state replacing the pseudonym by a personal identifier.

See

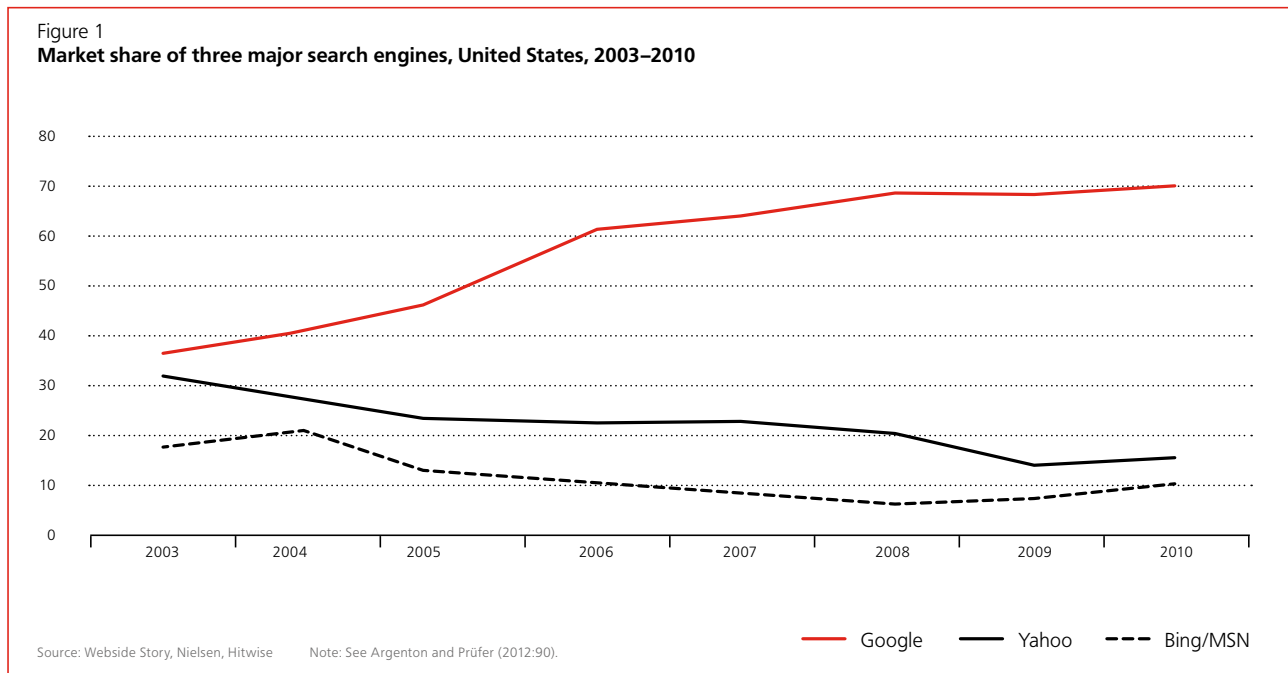
<https://www.protegrity.com/blog/pseudonymization-vs-anonymizationhelp-gdpr> for more explanations.

### 3.4 WHO EXACTLY SHOULD SHARE DATA?

According to the theoretical proposal of Prüfer and Schottmüller (2017), all firms in a data-driven market should share their user information with others. This seems suboptimal in practice, for at least two reasons: (i) data sharing comes at a cost and creates an administrative burden; (ii) large, dominant 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 new user information received (He et al., 2017). As the goal of the policy proposal is to establish, as much as possible, a contestable level playing-field, this suggests that large firms should share more data than small firms.

Based on this argumentation, the extremes are clear: the dominant firm must share all of its user information, whereas small firms do not have to share anything. The difficult task is to find the optimal boundary beyond which data should be shared. In principle, if only the firm serving most users is subject to a data-sharing obligation and if the second »largest« firm is not too far behind – for example, because a market has not tipped yet – it may give an unfair advantage to the second firm if it does not have to share its own user information. Together with the data from the largest firm, it may then have more user information in total than the leading one, and overtake the market leader, increasing its profits significantly, just because of an asymmetric data-sharing obligation.<sup>18</sup>

Therefore, it seems reasonable to introduce a threshold such that the largest two or three firms have to share data, as these are the contenders for »dominant firm« status. »Size«



would be measured by the amount of user information a provider has access to, which may be proxied by the number of individual interactions with users in most industries. A supervisory authority could collect such information relatively easily, at least for providers offering services connected to the internet and if providers have a duty to collaborate. An alternative to just determining that the »largest« two or three providers are obliged to share would be to set a market-share threshold such that only providers with market shares larger than 20 or 30 per cent have to share.

These approaches can also be combined: Oblige a firm to share its user information if it is among the top three firms in the industry and if it serves at least 20 per cent of the industry's users.

By way of example, Figure 1, which is taken from Argenton and Prüfer (2012), displays the development of market shares in the US search engine industry from 2003 to 2010. If the rule suggested above had been applied, MSN would have had to share its data only in 2004, Yahoo would have had to share between 2003 and 2008, and Google would have had to share throughout the entire period. No other search engine

**20** Mayer-Schönberger and Ramge (2018) propose to require that all firms with a market share above 10 per cent to share data – and to let the amount of data to be shared increase with the firm's market share (progressive data-sharing mandate). The ingenuity of this idea is that it introduces asymmetric sharing obligations, which should help to fight market tipping. The difficulty of the approach, however, is that it requires very precise estimates of market shares, which also change over time. Moreover, it introduces additional enforcement problems if, for instance, one provider has to share 40 per cent of its user information. How can it be ensured that the 40 per cent of the data that is shared is an unbiased and representative sample of that provider's entire user information? These problems are not impossible to solve but seem to be unnecessarily cumbersome. My preferred proposal also relies on »market shares« (measured by the amount of user information) to determine which firms are obliged to share data. Then, however, I propose to let them share all their user information, which gets rid of the sample problems

providers would have had to share anything. This asymmetric treatment might have avoided market tipping.<sup>21</sup>

If government agencies compete with private firms – that is, if they offer services in areas that are not particularly sensitive or security-related – they would be subject to the same sharing obligations.<sup>22</sup> The rationale is simply that user information is a virtually free by-product of running a service. Giving third parties access to those data is efficient as it enables the others to innovate and to compete with the incumbent, which benefits users.

**21** This example shows the mechanics of the threshold proposal: if a market has already tipped, only the dominant firm will have more than a 20 per cent market share. Hence, only one firm will have to share. If the market has not tipped yet, one may ask how the agency in charge of applying the test may determine whether the market is data-driven in the first place. To answer this question, the test for circularity (user information → quality → users → user information) described in Section 3.1 applies.

**22** By the logic of the proposal, this includes applications in which a government agency receives user information through logging interactions with citizens electronically but does not have the capacity to offer a valuable service. For instance, if a city government offers citizens with security concerns an app to communicate quickly in case of suspicious activities in their neighborhood and the received user information might improve that service (and the city government is dominant), a competing entrepreneur should have the right to obtain that (anonymized) user information, too. However, the proposal excludes areas in which the public agency obtains information about citizens because they were legally obliged to provide it, the information was processed at positive marginal cost and it was not obtained as a free by-product of machine-aided interaction between a citizen and the government agency (for example, during new registration in the city).

### 3.5 WHO SHOULD HAVE THE RIGHT TO OBTAIN ACCESS TO THE SHARED DATA? AT WHAT PRICE?

The guiding principle for answering this question is, yet again, the insight that user information is a free by-product of running a service. Some have claimed that because obtaining access to more user information helps to improve the quality of one's service, accumulating it is justified as an end in itself. However, the data gathered in this way is certain to be transformed into revenue at some point (for instance, through advertising or some other sale of access to one's user groups) and protecting those indirect revenues is not a goal of the policy proposal at hand because, in the long run, they are subject to the main market-tipping dynamics characterized by Prüfer and Schottmüller (2017). User information therefore has the attributes of a public good. 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, I propose that 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. For example, if a start-up wants to develop a service offering restaurant evaluations in Germany, then, based on the user-centred procedure suggested in Section 3.1, one would first have to determine which other firms offer such services. This may include specialists (for example, Yelp) but also general services (for example, Google Search/Maps). The start-up could demand user information from these providers (if they are dominant) that is relevant to serving consumers in the restaurant-evaluation business in Germany (but it could not ask Google for all its other user information!).

Inspiration can be drawn from the Payment Services Directive 2 (PSD2) in the financial industry, which entitles third parties, with the consent of the account holder, to access payment accounts in order to initiate payment transactions via an internet application or to consolidate account information from one or more accounts into one application. As such, PSD2 is a perfect example of EU regulation to level the playing field in the financial sector. Now banks have to accommodate fintech start-ups offering innovative services. The European Banking Authority has adopted several Guidelines and Regulatory Technical Standards related to implementing access to accounts, clarifying the steps banks need to take. Several initiatives are defining common API standards.

As on the sharing-obligation side, following the guideline that user information has public-good characteristics and was obtained virtually for free, the right to obtain shared user information pertains to for-profit, non-profit and public organizations (state authorities), subject to the above-mentioned condition of being able to demonstrate that they can and will offer a service to users utilizing the data. This guideline also indicates what the appropriate access price to another provider's user information should be, namely equal to the sharing provider's marginal cost of obtaining the user information,

which is of course zero.<sup>23</sup> Note that it is possible that some fixed costs may be associated with the storing of user information and the supervision of the data-sharing process. These costs will be negligible, however, compared with the costs of offering a service.<sup>24</sup> In any case, in order to make market entry and operations as easy as possible for the (small) competitors of dominant firms, they should not bear these costs.

### 3.6 HOW SHOULD DATA SHARING BE ORGANIZED? WHAT IS THE OPTIMAL GOVERNANCE STRUCTURE?

Turning to the question of the practical delineation of data-driven markets, addressed in Section 3.1, the issue of the optimal governance structure of data sharing is the one that needs most thought and research in the future. Here are a few initial thoughts that may inspire such analyses.

In institutional economics, there is a methodology that endogenizes optimal economic governance and corporate governance structures that tackle problems such as property rights protection, contract enforcement, or collective action.<sup>25</sup> This methodology can also be applied to mandatory data sharing in data-driven markets. Key dimensions of relevant economic governance institutions are centralized versus decentralized enforcement (of data sharing), public versus private enforcement, and coercive versus ostracism-based enforcement. In line with Section 3.1, it must be underlined that the optimal governance structure may differ between industries, depending on attributes such as the number of players, the extent to which a specific industry is national or international, the speed of technological progress (which can be assumed to be fairly high in all data-driven markets), and the homogeneity of interests of senders and receivers of user information.

Section 3.4 proposes that no more than three firms per industry be subject to a data-sharing obligation. Section 3.5, however, suggests that many organizations might have a valid claim to the shared data. Therefore, decentralized data sharing, that is, direct interconnections between senders and re-

<sup>23</sup> Note that this result is very different from access pricing in utility industries such as telecommunications, internet backbone or railways because there the marginal cost for access to the shared resource (for example, utilization of a telecommunications network to offer one's own services) is positive. Additionally, there are positive fixed costs of operation, for example, for maintenance. On data-driven markets, by contrast, there are virtually no costs involved in collecting user information (only in running the main service, for example, internet search or restaurant evaluations).

<sup>24</sup> For instance, US data centers have consumed about 2% of the country's entire energy consumption in 2014 (<https://www.datacenterknowledge.com/archives/2016/06/27/heres-how-much-energy-all-us-data-centers-consume>). Google's consumption alone has nearly doubled 2014-2018 (<https://www.statista.com/statistics/788540/energy-consumption-of-google/>). These numbers underline the massive scale at which data-driven markets affect the offline world already now --- and which any entrepreneur would have to compete with.

<sup>25</sup> See Dixit (2009) and Williamson (2005) for general introductions and Prüfer (2013 and 2018) for applications of this methodology to the problem of a lack of trust in cloud computing.



ceivers (for example, via APIs) seems impractical. By contrast, given that there may be hundreds of sender–receiver relations per data-driven industry and that all should be able to compete on a level playing-field, an intermediary is needed between senders and receivers (implementing centralized enforcement). This notion is strengthened if one revisits Section 3.3, on privacy and recalls that the shared data has to be anonymized before it reaches the receiver, for instance, in a data pool of user information that is only accessed by algorithms, for training purposes, but never by human beings. This could occur in a standardized fashion at the stage of the intermediary and not in a decentralized way at the stage of the sharing providers.

Many global high-tech markets can be well governed by private organizations, such as industry associations, which have a lot of know-how, often involve effective arbitration tribunals and can react quickly to changing technological, economic, political, or legal circumstances (see Prüfer, 2013, 2018). A peculiarity of data-driven markets is, however, that the interests of the dominant firm and those of all other firms are opposed: while all other providers want quick, reliable and comprehensive data sharing, the dominant firm wants to keep its exclusive advantage in access to user information and hence has high incentives to obstruct data sharing. Therefore, it seems optimal to commission a public organization with access to the coercive enforcement powers of the state to manage, or at least closely monitor, data sharing.

This intermediary organization would be tasked with the structure and operation of the data-sharing scheme. It would have to cooperate closely with the data-sending firms in an industry, to validate the business plans of would-be receivers and to make sure that all certified receivers receive the appropriately anonymized user information of senders in a standardized, equitable and workable way. To reap economies of scope, it would also be tasked with market definition and the running of tests on the data-drivenness of all potential industries (see Section 3.1).

Given this massive volume of important tasks, the organization must be well equipped with resources, especially with experts from various domains (at least technological, legal and economic) and with appropriate powers to perform its tasks effectively. Moreover, the organization must have a legitimate mandate to perform its tasks throughout the European Union. Therefore, it seems natural to locate it at EU level. It could be a new part of the EU's DG Competition or an independent agency with appropriate resource endowments and powers.<sup>26</sup> Moreover, the organization – let us call it the European Data Sharing Authority for now – should collaborate with national authorities in member states, especially in cases in which markets, despite their internet-related character, are largely local (for example, platforms for food delivery).

---

**26** One legal scholar has expressed the following view: »I think in any case this needs to be run through a legislative process: either by adopting frameworks for specific sectors like PSD2 or by granting competition authorities clearly defined new competences. Because they are executive bodies, competition authorities in my view lack the legitimacy to make such important policy choices on the basis of the current competition law framework.«

## 4

## CONCLUSION

In line with rapid technological progress in the fields of data collection (creating big data sets) and data analytics (driven by machine learning), academic research in law and economics has started to recognize and analyse the problems stemming from these developments over the past few years. Since Argenton and Prüfer (2012), the identification of market tipping as a central consequence of datafication and the policy proposal of mandatory data sharing has gained more and more traction, both among researchers and, since 2016, among policy-makers. Moreover, all large and many smaller countries have commissioned expert reports on the economics of »platform markets«, »digital markets« and »big tech regulation«, and most of them have concluded that existing competition law has to be developed further to keep up with the challenges in such industries.<sup>27</sup> Many include a provision that would allow competition authorities to mandate data sharing in specific industries, especially if the network effects driving industry dynamics are very strong.<sup>28</sup>

In the academic literature I am aware of, no other policy proposal is as concrete as the mandatory data-sharing proposal to fight the ills of market tipping on data-driven markets. Therefore, in this article I have attempted to develop this proposal further and to bring it closer to legal implementability by suggesting tentative answers to the key questions that have come up in many discussions with policy-makers, practitioners and academics.

As already explained, more research is needed. Most urgent is the development of an empirical test for data-drivenness at the industry level and the application of the test in the field (see Section 3.1). The second most pressing issue is to study the economic and corporate governance of data sharing in order to better understand how the proposal could be imple-

mented in a way that gets the incentives of all involved parties right (see Section 3.6).

One important caveat concerning the proposal is that, even if mandatory data sharing was fully implemented in the EU, as outlined above, it is only a necessary, not a sufficient condition for lively competition and innovation on data-driven markets. Competing successfully with the most valuable and technologically advanced firms on the planet in the markets they have dominated for years, requires more than access to the raw data those firms collect by logging interactions with users. Would-be competitors will also depend on highly skilled and specialized workers, especially those who understand today's AI algorithms, and first-class access to assets such as infrastructure, software and hardware.<sup>29</sup> After data sharing is implemented, in these complementary dimensions, which are beyond the control of most single firms, policy-makers may play a role.

Even if this may appear to be a Herculean task, all efforts would be certainly lost without access to dominant firms' user information. Only if entrepreneurs and other would-be competitors can benefit from free access to information about users' preferences and characteristics, which is essential in data-driven markets, will it be worth it for them to develop new products, services and quality improvements and for financial investors to fund their ventures. The sheer prospect of such competition, in turn, will motivate the dominant firms to fight for their turf and to innovate harder themselves. The result will be that users benefit from more choice, both between and within providers, from higher product and service quality, and lower prices. Dominant firms will have less market power, which at present gives them sufficient leeway to engage in anticompetitive conduct (whether or not they use this leeway). Finally, among emerging competitors a unicorn may come to the fore that gets its chance to flourish by means of access to user information today, which would otherwise have perished.

<sup>27</sup> Examples include the Vestager Report in the EU (Crémer et al., 2019), the Furman Report in the UK (Furman et al., 2019), the Stigler Committee on Digital Platforms in the US (Zingales et al., 2019) or the Bericht der Kommission Wettbewerbsrecht 4.0 (Schallbruch et al., 2019) in Germany. Beaton-Wells (2019:2) writes of the report by the Australian Competition and Consumer Commission »there have been just shy of 30 such inquiries on the same or related topics published or announced around the world in the last five years«.

<sup>28</sup> For instance, the "Dutch vision on data sharing between businesses" (2019) or Soriano (2019), the Head of the French regulator ARCEP, go far in this direction.

<sup>29</sup> In many data-driven markets, the nominal price to use the services of a provider are zero. However, often the economic price paid is expressed in a different currency such as one's attention, personal data, or the agreement to terms & conditions one might have rejected otherwise. See the recent Facebook case at Germany's Bundeskartellamt. Dengler and Prüfer (2018) show how the option to escape personalized prices through using an anonymous sales channel can lead to net benefits for some but to losses for other users.





# References

- Agrawal, Ajay, Gans, Joshua; Avi Goldfarb** 2018: *Prediction Machines: The Simple Economics of Artificial Intelligence*, Boston.
- Argenton, C.; Prüfer, J.** 2012: Search Engine Competition with Network Externalities, in: *Journal of Competition Law and Economics*, S. 73–105.
- Autor, David; Dorn, David; Katz, Lawrence F.; Patterson, Christina; John Van Reenen** 2019: The Fall of the Labor Share and the Rise of Superstar Firms, NBER Working Paper No. 23396.
- Beaton-Wells, Caron** 2019: Ten Things to Know About the ACCC's Digital Platforms Inquiry, Competition Policy International, Melbourne.
- Brandom, Russel** 2018: Google, Facebook, Microsoft, and Twitter Partner for Ambitious New Data Project: An Open-Source Collaboration for the Future of Portability, <https://www.theverge.com/2018/7/20/17589246/data-transfer-project-google-facebook-microsoft-twitter> (12.1.2020).
- Bughin, J.; Seong, J.; Manyika, J.; Chui, M.; R. Joshi** 2018: Notes from the AI Frontier: Modeling the Impact of AI on the World Economy, Discussion Paper McKinsey Global Institute.
- Crémer, J.; Montjoye, Y.-A.; H. Schweitzer** 2019: Competition Policy for the Digital Era, Report for European Commission, DG Competition, <https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf> (12.1.2020).
- Dengler, S.; J. Prüfer** 2018: Consumers' Privacy Choices in the Era of Big Data, TILEC Discussion Paper No. 2018-014.
- Dixit, A. K.** 2009: Governance Institutions and Economic Activity, in: *American Economic Review* 99 (1), S. 5–24.
- Ferschli, Benjamin; Rehm, Miriam; Schnetzer, Matthias; Zilian, Stella** 2019: *Marktmacht, Finanzialisierung, Ungleichheit: Wie die Digitalisierung die deutsche Wirtschaft verändert*, Friedrich-Ebert-Stiftung, Bonn, <http://library.fes.de/pdf-files/fes/15744.pdf> (31.1.2020).
- European Commission** 2019: Antitrust/Cartel Cases, [http://ec.europa.eu/competition/elojade/isef/case\\_details.cfm?proc\\_code=1\\_39740](http://ec.europa.eu/competition/elojade/isef/case_details.cfm?proc_code=1_39740) (12.1.2020).
- Fezer, Karl-Heinz** 2018: *Repräsentatives Dateneigentum: Ein zivilgesellschaftliches Bürgerrecht*, Studie im Auftrag der Konrad-Adenauer-Stiftung.
- Financial Times** 2019: How Top Health Websites Are Sharing Sensitive Data with Advertisers, 13.11.2019.
- Furman, J.; Coyle, D.; Fletcher, A.; McAuley, D.; P. Marsden** 2019: Unlocking Digital Competition, Report of the (UK) Digital Competition Expert Panel, [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/785547/unlocking\\_digital\\_competition\\_furman\\_review\\_web.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf) (12.1.2020).
- Gebicka, A.; Heinemann, A.** 2014: Social Media & Competition Law, in: *World Competition* 37 (2), S. 149–172.
- Government of the Netherlands** 2019: Zukunftstauglichkeit des Wettbewerbsinstrumentariums in Bezug auf Onlineplattformen, <https://www.government.nl/documents/letters/2019/05/23/zukunftstauglichkeit-des-wettbewerbsinstrumentariums-in-bezug-auf-onlineplattformen> (12.1.2020).
- Graef, I.; Husovec, M.; Purtova, N.** 2018: Data Portability and Data Control: Lessons for an Emerging Concept in EU Law, in: *German Law Journal*, 19 (6), S. 1.359–1.398.
- He, D.; Kannan, A.; McAfee, R. P.; Liu, T.-Y.; Qin, T.; Rao, J. M.** 2017: Scale Effects in Web Search, in: Devanur, N. R.; P. Lu (Hrsg.): *WINE 2017*, LNCS 10674, [https://doi.org/10.1007/978-3-319-71924-5\\_21](https://doi.org/10.1007/978-3-319-71924-5_21).
- LeCun, Y.; Bengio, Y.; G. Hinton** 2015: Deep Learning, *Nature* 521, <https://www.cs.toronto.edu/~hinton/absps/NatureDeepReview.pdf> (12.1.2020).
- Lee, MJ; DePillis, Lydia; Krieg, Gregory** 2019: Elizabeth Warren's New Plan: Break up Amazon, Google and Facebook, in: *CNN*, 8.3.2019, <https://edition.cnn.com/2019/03/08/politics/elizabeth-warren-amazon-google-facebook/index.html> (12.1.2020).
- Loechner, Jack** 2016: 90 % of Today's Data Created in Two Years, in: *Media Post*, 22.12.2016, <https://www.mediapost.com/publications/article/291358/90-of-to-days-data-created-in-two-years.html> (30.1.2020).
- Mayer-Schönberger, V.; T. Ramge** 2018: *Reinventing Capitalism in the Age of Big Data*, London.
- Ministry of Economic Affairs and Climate Policy (of the Netherlands)** 2019: Dutch Vision on Data Sharing Between Businesses, <https://www.government.nl/documents/reports/2019/02/01/dutch-vision-on-data-sharing-between-businesses> (12.1.2020).
- New York Times** 2019: Google to Store and Analyze Millions of Health Records, 11.11.2019.
- Page, L.; Brin, S.; Motwani, R.; T. Winograd** 1998: The Pagerank Citation Ranking: Bringing Order to the Web, Technical Report, Stanford.
- Prüfer, J.** 2013: How to Govern the Cloud?, *IEEE CloudCom 2013*, DOI 10.1109/CloudCom.2013.100, S. 33–38.
- Prüfer, J.** 2018: Trusting Privacy in the Cloud, *Information Economics and Policy* 45, S. 52–67.
- Prüfer, J.; Schottmüller, C.** 2017: Competing with Big Data, TILEC Discussion Paper No. 2017–006, Tilburg.
- Schaefer, M.; Sapi, G.; S. Lorincz** 2018: The Effect of Big Data on Recommendation Quality: The Example of Internet Search, *DIW Discussion Paper 1730*, [http://www.dice.hhu.de/fileadmin/redaktion/Fakultaeten/Wirtschaftswissenschaftliche\\_Fakultaet/DICE/Discussion\\_Paper/284\\_Schaefer\\_Sapi\\_Lorincz.pdf](http://www.dice.hhu.de/fileadmin/redaktion/Fakultaeten/Wirtschaftswissenschaftliche_Fakultaet/DICE/Discussion_Paper/284_Schaefer_Sapi_Lorincz.pdf) (12.1.2020).
- Schallbruch, M.; Schweitzer, H.; Wambach, A.** (Hrsg.): *Ein neuer Wettbewerbsrahmen für die Digitalwirtschaft*, Bericht der Kommission Wettbewerbsrecht 4.0. BMWi, Berlin.
- Schultz, Jeff** 2019: How Much Data is Created on the Internet Each Day?, in: *Micro Focus Blog*, 8.6.2019, <https://blog.microfocus.com/how-much-data-is-created-on-the-internet-each-day/#> (12.1.2020).
- Soriano, S.** 2019: Big Tech Regulation: Empowering the Many by Regulating a Few, [https://www.arcep.fr/fileadmin/reprise/communiqués/communiqués/2019/pdf/Big-Tech-Regulation\\_MediumSSo-avril2019.pdf](https://www.arcep.fr/fileadmin/reprise/communiqués/communiqués/2019/pdf/Big-Tech-Regulation_MediumSSo-avril2019.pdf) (12.1.2020).
- SPD** (2019a): Digitaler Fortschritt durch ein Daten-für-Alle-Gesetz, [https://www.spd.de/fileadmin/Dokumente/Sonstiges/Daten\\_fuer\\_Alle.pdf](https://www.spd.de/fileadmin/Dokumente/Sonstiges/Daten_fuer_Alle.pdf) (12.1.2020).
- SPD** (2019b): Beschlüsse & Anträge, 06.-08.12.2019, <https://indieneuzeit.spd.de/beschluesse/> (11.2.2020).
- Statcounter** o. J.a: Search Engine Market Share Worldwide, <https://gs.statcounter.com/search-engine-market-share/all> (31.1.2020).
- Statcounter** o. J.b: Social Media Stats Europe, <https://gs.statcounter.com/social-media-stats/all/europe> (31.1.2020).
- Sverdlík, Yevgeniy** 2016: Here's How Much Energy All US Data Centers Consume, <https://www.datacenterknowledge.com/archives/2016/06/27/heres-how-much-energy-all-us-data-centers-consume> (12.1.2020).
- University of St. Gallen** o. J.: How Amazon dominates the market in Germany, <https://item.unisg.ch/en/news/amazon-watch-report-1> (31.1.2020).
- Wang, T.** 2019: Energy Consumption of Google from 2011 to 2018, in: *Statista*, 10.10.2019, <https://www.statista.com/statistics/788540/energy-consumption-of-google/> (12.1.2020).
- Williamson, Oliver E.** 2005: The Economics of Governance, *American Economic Review* P&P, 95 (2), S. 1–18.
- Zingales, L.; Rolnik, G.; Lancieri, F.M.** (Hrsg.): *Stigler Committee on Digital Platforms*, Final Report, Chicago.



## The Friedrich-Ebert-Stiftung

The Friedrich-Ebert-Stiftung (FES) is the oldest political foundation in Germany with a rich tradition dating back to its foundation in 1925. Today, it remains loyal to the legacy of its namesake and campaigns for the core ideas and values of social democracy: freedom, justice and solidarity. It has a close connection to social democracy and free trade unions.

FES promotes the advancement of social democracy, in particular by:

- political educational work to strengthen civil society;
- think tanks;
- international cooperation with our international network of offices in more than 100 countries;
- support for talented young people;
- maintaining the collective memory of Social Democracy with archives, libraries and more.

## IMPRINT

© 2020

### **Friedrich-Ebert-Stiftung**

Godesbeger Allee 149, D-53175 Bonn

Orders/contact: [BeMo@fes.de](mailto:BeMo@fes.de)

The views expressed in this publication are not necessarily those of the Friedrich-Ebert-Stiftung. The commercial exploitation of the media published by the FES is allowed only with the written permission of the FES.

**ISBN: 978-3-96250-527-1**

Picture: © Adobe Stock/garrykillian

Design concept: [www.bergsee-blau.de](http://www.bergsee-blau.de)

Print: [www.bub-bonn.de](http://www.bub-bonn.de)

## FOR A BETTER TOMORROW

### Competition Policy and Data sharing on Data-driven Markets. Steps Towards Legal Implementation.

The big tech companies continue to grow and increasingly dominate the commercial internet. Over 90 per cent of internet searches go through Google; in social media Facebook's European market share is over 70 per cent; and almost half of Germany's online commerce now takes place via Amazon. This tendency towards the monopolisation of data-driven markets not only endangers competition, but weakens business's innovative capacity.

Regulators on both sides of the Atlantic are now looking at what can be done to counteract this at the political level. Most proposals rely on traditional competition-law instruments. The present study, by contrast, makes a far-reaching recommendation. The basic assumption is that the massive accumulation of data in data-driven markets and the denial of access to it represent the main causes of increasing monopolisation. The decisive measure to break through the tech monopoly, then, would be to open up data to competitors. Only an obligation to share data (in conformity with data protection) could ensure competition and innovation in data-driven markets, according to the author.

### The Author

**Jens Prüfer** is Associate Professor of Economics at Tilburg University and a member of the Tilburg Law and Economics Center (TILEC).



For further information please visit:  
[www.fes.de/fuer-ein-besseres-morgen](http://www.fes.de/fuer-ein-besseres-morgen)