THE COMPETITIVE EFFECTS OF THE GDPR

*Michal S. Gal and Oshrit Aviv**

Forthcoming, *Journal of Competition Law and Economics* (2020)

TABLE OF CONTENTS

I. INTRODUCTION 1
II. SETTING THE STAGE 7
A. Three Assumptions Regarding Data Markets *7*
B. Five Potential Business Models for Obtaining Data *9*
III. SHAPING CHOICES: THE POTENTIAL EFFECTS OF LEGAL LIMITATIONS ON THE FIVE BUSINESS MODELS 11
IV. POTENTIAL LEGAL AND TECHNOLOGICAL BYPASSES 25
A. Legal Obligations to Share Data *25*
B. Technological Solutions? 27
V. POTENTIAL EFFECTS ON COMPETITION AND WELFARE 28
VI. CONCLUSION AND THE WAY FORWARD 33

"To share data or not to share, that is the question..."

I. INTRODUCTION

In his famous picture, The Ambassadors, Hans Holbein the Yong painted a skull, a symbol of mortality. This skull is a well-known example of anamorphosis: it can only be seen when looking at the picture from a certain perspective.[1] In this paper, we engage in legal anamorphosis, analyzing the effects of the General Data Protection Regulation (GDPR)[2] from a perspective which to a large extent remains unexplored—namely, its competitive effects. Our aim here is to provide a more holistic view of the effects of the GDPR on social welfare, and to incentivize regulators to take steps to ensure that government-mandated data governance tools indeed increase social welfare.

---

* Professor of Law and Director of the Center for Law and Technology, University of Haifa Faculty of Law, and President of the International Association of Competition Law Scholars (ASCOLA); Lawyer, Information System Scientist, Data Privacy Specialist and CEO at Entero.io, a data and technology compliance-by-design company. Many thanks to Stefan Bechtold, Damien Geradin, Inge Graef, Vikas Kathuria, Giorgio Monti, Tomas Thombal, Nicolo Zingales, and participants in the workshop 'Governing Data as a Resource' organised by the Tilburg Institute for Law, Technology, and Society (TILT) and the Tilburg Law and Economics Center (TILEC) on 22 November 2019, for most helpful comments on previous drafts, and to Tamar Shmueli for valuable research assistance. The authors thank the Center for Cyber Law and Policy at the University of Haifa for funding. Any mistakes or omissions remain the authors'.

[1] Hans Holbein the Younger, The Ambassadors, 1533, National Gallery, London.

[2] Regulation (EU) 2016/679 of the European Parliament and of the Council 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, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L 119/1, 4 May 2016.

The increased importance of data in our information economy as an input for innovation, economic growth, and societal interactions, along with the associated surveillance and security concerns,[3] have brought data collection, processing, and use issues to the fore. Such issues are affected by a combination of private incentives and public measures. For example, voluntary sharing of data may be constrained by legal limitations.[4] The converse is also true: a data controller's unwillingness to share his data might be overridden by legal data sharing obligations.[5] Understanding the interaction between private incentives and regulatory measures is thus essential for designing data governance models that can enable a well-functioning and social-welfare-enhancing digital economy.

This paper focuses on one important aspect of this public/private interaction, which so far has been largely neglected: an in-depth analysis of the effects of the GDPR on competition and innovation, in particular the ability to realize economies in scale and scope in data analysis. Commentators have recognized that the high costs of putting in place a GDPR-compliant system might advantage large firms,[6] that consent-based data collection creates comparative advantages to diversified or large firms which collect their own data,[7] that the use of some methods to collect data, such as third-party tracking, has decreased,[8] and that the costs of ensuring that an acquisition target is GDPR-compliant might reduce incentives to merge.[9] Empirical studies have pointed to potential negative effects of the GDPR on competition and investment.[10] Still, to our knowledge, no in-depth systematic analysis of the competitive effects of the legal provisions of the GDPR has yet been undertaken. This is the task we attempt here.

---

[3] The Snowden revelations and similar privacy breaches generated a response in the form of another layer of regulations covering the collection and use of data, and justifications for such.

[4] As elaborated in this article, the GDPR is a prime example. Article 4(2) of the GDPR defines data processing to include "making data available."

[5] See discussion *infra*.

[6] See, e.g., i-Scoop, *How the GDPR Impacts and Suffocates Small and Medium Businesses*, https://www.i-scoop.eu/gdpr/gdpr-small-medium-businesses/; Yedidiha Yueh, *GDPR Will Make Big Tech Even Bigger*, FORBES (June 26, 2018), https://www.forbes.com/sites/forbestechcouncil/2018/06/26/gdpr-will-make-big-tech-even-bigger/#41977f6e2592.

[7] James David Campbell, Avi Goldfarb, and Catherine E. Tucker, *Privacy Regulation and Market Structure*, 24(1) JOURNAL OF ECONOMICS & MANAGEMENT STRATEGY 47 (2015).

[8] P. Wagner, *News Pages Are Abandoning Third-Party Ad Trackers*, September 25th, 2018, https://www.statista.com/chart/15578/change-of-ad-tracking-techniques-since-gdpr/ (since the GDPR came into force third party cookies decreased by 22 percent per page and third-party domains decreased by 4 percent).

[9] See, e.g., Erin Grant, *The Crucial Implications of GDPR on Mergers and Acquisitions*, congruity360 14.6.2018; Yahoo News, *A new survey shows how the GDPR is impacting mergers and acquisitions*, 26.11.2018, https://finance.yahoo.com/news/survey-shows-gdpr-impacting-mergers-160019175.html; Tim Bird, *GDPR: Managing Privacy is Now Key to Mergers and Acquisitions*, PERSONNEL TODAY, May 24, 2019.

[10] For a combined source of such studies see, e.g., Center for Data Innovation, Eline Chivot and Daniel Castro, *What the Evidence Shows About the Impact of the GDPR After One Year*, June 17, 2019. We do not regard the content of the empirical studies referred to throughout this paper as necessarily accurate. *See also* Michail Batikas et al., *European Privacy Law and Global Markets for Data* (2020, on file with author)(finding an increase in market concentration in web technology services after the introduction of the GDPR. Google, the dominant firm in many markets for web technologies, manages to increase its market share, whereas all other firms that supply web technology either do not see a change in market share or suffer losses.), and sources cited there.

The importance of the GDPR cannot be overstated.[11] It seeks to protect consumers and users from harms resulting from unauthorized and excessive use of their personal data (hereinafter, the term data will refer to personal data,[12] unless specified otherwise), in ways that might negatively affect human dignity and well-being, including but not limited to price discrimination, other forms of discrimination, blackmail, intangible nuisances, identity theft, and harm to autonomy.[13] The GDPR also seeks to change the balance of power between data subjects and data controllers,[14] potentially enabling data subjects to enjoy a larger portion of the fruits from sharing their data. Additionally, it seeks to ensure the free flow of data between EU member states, inter alia by eliminating differences among such states with regard to data processing.[15] Furthermore, it seeks to strengthen the trust users have that their personal data will not be used in ways that do not comport with their reasonable expectations, which is necessary for the efficient working of the market and for society to realize the value of technology.[16]

At the same time, the GDPR creates inherent tradeoffs between data protection and other dimensions of welfare, including competition and innovation.[17] While some of these effects were acknowledged when constructing the legal data regime, many were disregarded.[18] Furthermore, the magnitude and breadth of such effects may well constitute an unintended and unheeded welfare-reducing consequence.[19] As this article shows, the GDPR limits competition and increases concentration in data and data-

[11] The right to the protection of personal data is enshrined in Article 8 of the EU Charter of Fundamental Rights and in Article 16(1) of the Treaty on the Functioning of the European Union. In this article we assume that the GDPR effectively protects privacy. Some question the proposition that a system which relies mainly on notice and consent as a basis for the lawful collection and use of data can indeed protect privacy. See, e.g., Written Statement For The Record David Hoffman, Associate General Counsel and Global Privacy Officer, Intel Corporation, United States Senate Committee on the Judiciary Hearing on "GDPR & CCPA: Opt-ins, Consumer Control, and the Impact on Competition and Innovation" (12.3.2019)(" The notice-and-consent model is fatally flawed; it must be replaced. People do not have time to read privacy policies for every interaction where their personal data will be collected and used. Even if they did read these policies, it is unlikely they would be able to understand how this data will be used"); Written Statement For The Record, Jane Bambauer, *ibid* ("A GDPR-style of privacy right that gives consumers and end users full control over personal information has enormous popular appeal, but ...will burden the digital economy with transaction costs, and there is little reason to think that the compliance costs or behavioral changes will have a meaningful relationship to harm"); Nickolas Economides and Ioannis Lianos, *Restrictions on Privacy and Exploitation in the Digital Economy: A Competition Law Perspective* (2019), https://www.ucl.ac.uk/cles/sites/cles/files/cles-5-2019.pdf (a market failure exist with regard to privacy protection)

[12] For the definition of personal data and its boundaries see Article 4(a) of the GDPR.

[13] Alessandro Acquisiti, Curtis Taylor, and Liad Wagman, *The Economics of Privacy*, 54(2) JOURNAL OF ECONOMIC LITERATURE 442 (2016); Niva Elkin-Koren and Michal S. Gal, *The Chilling Effects of Governance by Data*, 86 CHI. L. REV. (2019).

[14] This term shall relate to data controllers and processors, as defined in the GDPR, unless specified otherwise.

[15] Recitals 3 and 9 of the GDPR.

[16] For the effect of trust on markets see, e.g., Elkin-Koren and Gal, *supra* note 13.

[17] For an example in the healthcare industry see William Nicholson Price et al., *Shadow Health Records Meet New Data Privacy Laws*, 363(6426) SCIENCE 448 (2019)(the GDPR prevents the collections of health data outside the health system that provide detailed accounts of individual health, which can be used for innovation and quality measurements).

[18] See discussion *infra* of the Commission's Impact Assessment of the GDPR.

[19] Of course, the GDPR is not the only law which creates compliance costs which might affect competition. See, e.g., Miguel de la Mano and Jorge Padilla, *Big Tech Banking* (2018), available at https://ssrn.com/abstract=3294723. Accordingly, the effect might be cumulative.

related markets, and potentially strengthens large data controllers. It also further reinforces the already existing barriers to data sharing in the EU,[20] thereby potentially reducing data synergies that might result from combining different datasets controlled by separate entities.[21]

To illustrate its claims, the article analyzes the competitive dynamics created by the GDPR, focusing on how it affects the options available to firms for amassing the data necessary for their operations, and their resultant ability to realize economies of scale and scope in data analysis. A special focus is placed on the ability of firms to share the data with other firms. Our analysis is based on a careful evaluation of GDPR provisions and guidelines, as well as on interviews and survey data from firms for which data constitutes an important input.[22] Even at this early stage, in which market players are still evaluating the effects of the new legal data regime on their operations, the analysis reveals important effects on competition and innovation, many of which are long-term.

We identify seven main parallel and cumulative market dynamics that may limit competition and increase market concentration, of which only some have been recognized so far. First, as some commentators have already observed,[23] the costs of organizing a dataset in a way which complies with the GDPR may be high and are characterized by economies of scale. Accordingly, some small entrants might find it unprofitable to collect data. Second, also as previously observed, the GDPR prohibits or makes it more difficult to engage in some methods of data collection, creating comparative advantages to some data controllers. For example, in their seminal article Campbell, Goldfarb and Tucker showed that the need to receive a user's consent to use his data imposes transaction costs for internal data collection, whose effects fall disproportionately on less diversified or new firms.[24] Both dynamics reduce the number of potential competitors in data collection.

Third, the GDPR reduces the economic incentives of firms to share any data collected. This is because those sharing data are still liable for monitoring its use by anyone with whom the data is shared (hereinafter: data receivers).[25] This, in turn, further reduces the number of data suppliers.

Fourth, even where data is shared, the GDPR may limit its use. To illustrate, it is often costly, and sometimes impossible, to obtain informed consent from data subjects to have their data shared with the data receiver, as may be required by the GDPR. This effect is strengthened in a multi-product and/or multi-service environment, in

---

[20] IDC and Open Evidence, *European Data Market Study* (2017) 75 https://ec.europa.eu/digital-single-market/en/news/final-results-european-data-market-study-measuring-size-and-trends-eu-data-economy. Only 6.3% of European undertakings reported that they took an active part in data sharing and reuse.

[21] *See, e.g.*, OECD, DATA-DRIVEN INNOVATION: BIG DATA FOR GROWTH AND WELL-BEING (2015) at 193 (describing how the value of data "increases when the data can be linked with and integrated into other data sets").

[22] The questionnaire used for the survey is attached as Appendix A.

[23] International Association of Privacy Professionals, IAPP-EY Annual Governance Report 2018 (2019), https://iapp.org/resources/article/iapp-ey-annual-governance-report-2018/ (the average firm of 500 employees must spend about $3 million to comply with the GDPR).

[24] Campbell *et al.*, *supra* note 7 ("The commonly used consent-based approach may disproportionately benefit firms that offer a larger scope of services...This negative effect is particularly severe for goods where the price mechanism does not mediate the effect, such as the advertising-supported internet"). The model assumes a one-time consent rather than a consent which grows with the use of the data. The GDPR imposes some consent obligations which increase in line with privacy risks.

[25] See discussion *infra*.

which consent is required for each different use of the data. The stronger the legal limitations on using data collected by an external entity,[26] the stronger the motivation to collect it internally.

Fifth, the costs of non-compliance are high. Commentators point to the size of fines that can be imposed on firms which fail to comply with the GDPR.[27] To this we add another effect which can sometimes be more significant: the virality of non-compliant data. The GDPR imposes a duty on the data receiver to ensure that any data received from an external entity is GDPR-compliant. Accordingly, should non-compliant data be transferred from an external data controller and combined with the receiver's data, the whole dataset could be polluted (i.e., considered non-compliant). Virality may affect all types of data included in the dataset, including non-personal data, so long as it is combined with—and cannot be easily separated from—the non-compliant personal data. Furthermore, and potentially more troubling, even if the datasets can be separated ex post, any learning by an algorithm based on the combined dataset cannot be easily reversed, especially if such learning was already translated into products or services. Undoing such effects could significantly disrupt business operations. To avoid such consequences, data receivers must engage in ongoing monitoring of their data suppliers' collection and processing practices. This, in turn, might further reduce incentives to use externally collected data, and strengthen incentives for internal data collection.

Sixth, the GDPR creates uncertainty, which may impose higher costs on smaller players, and might also enable large firms to use such uncertainty strategically, limiting the sharing of their data based on broad interpretations of the GDPR.[28] Finally, the GDPR, and especially the discussions surrounding it, could have an indirect effect on data subjects, who might be more willing to provide their data to larger, more reputable firms, or to firms with which they must interact, at least until the trust of data subjects in the actual enforcement of data protection obligations is increased.[29]

The cumulative effect of such dynamics, explored in detail below, is a decline in competition in data (and in data-based) markets. More often than in the pre-GDPR period, [30] firms may now prefer to collect data internally. Where internal collection is costly or impossible, firms will prefer to purchase data from external data suppliers. Yet the GDPR reduces the number of potential data suppliers and increases the costs of and barriers to data-sharing transactions. Accordingly, it is now substantially more difficult for firms to realize data synergies through data sharing. Furthermore, where data-based analysis requires a combination of personal and non-personal data, or

[26] In this article, an external entity does not include a data processor who simply performs tasks for the data controller.

[27] Organizations in breach of GDPR can be fined up to 4% of annual global turnover or €20 million, whichever is greater. Article 83 of the GDPR.

[28] See discussion *infra*.

[29] See discussion *infra*.

[30] While limitations on data collection, processing and use also existed under the Data Protection Directive 95/46EC, the GDPR amplified them, inter alia by increasing fines for non-compliance, setting stricter requirements for data subjects' consent, widening the territorial scope of the application and the definition of personal data, and mandating privacy-by-design and the appointment of data protection officers in some instances.

where difficulties arise in separating these two types of data,[31] the effects of such obligations may carry over to non-personal data. Thus, the GDPR might also indirectly affect the free flow of non-personal data.

These effects may harm productive and dynamic efficiency. They also call into question the ability of EU firms to employ data sharing so as to increase their international comparative advantages in data-based markets, and to limit the market power of those entities which already possess data-based comparative advantages.

The dynamics identified in this article offer partial explanations for some of the troubling empirical evidence regarding investment in data-driven markets following the adoption of the GDPR.[32] A study conducted by the Merrill Corporation, for example, found that 58% of mergers and acquisitions professionals surveyed reported having worked on transactions that did not go through due to concerns about the parties' compliance with the GDPR.[33] Another study found that, post-GDPR, the number of deals involving EU ventures with data-related business activities decreased by almost 31%.[34] Our study offers some potential explanations to the correlations found in such studies. Furthermore, by identifying the market dynamics that affect data collection and competition, our study enables us to differentiate between the short- and long-term competitive effects of the GDPR, and to identify which effects are likely to dissipate and which are here to stay.

The article proceeds as follows. To set the stage for our regulatory impact assessment, Chapter II analyzes three basic assumptions regarding data markets, and five alternative business models for gathering data: internal growth, merging, data-sharing joint ventures, purchasing data, or purchasing data-based knowledge. Chapter III then analyzes the main legal limitations imposed by the GDPR which affect the choice between the five different business models. Chapter IV explores whether legal data-sharing obligations or technological means can reduce the negative effects of the GDPR on competition. First, we analyze the legal duty of data sharing based on the right to data portability and the refusal-to-deal prohibition, to determine whether, and to what extent, such mandatory requirements affect the dynamics of data sharing. We then analyze the extent to which available technological solutions can assist firms in overcoming some of the legal barriers and costs created by the GDPR. Chapter V

---

[31] Such difficulties might be technical, such as identifying which data are private, or they may be legal, such as defining what constitutes private data under the law. *See, e.g.*, Inge Graef *et al*., *Feedback to the Commission's Proposal on a Framework for the Free Flow of Non-Personal Data* (2018), ssrn.com/abstract=3106791.

[32] Center for Data Innovation, Eline Chivot and Daniel Castro, *What the Evidence Shows About the Impact of the GDPR After One Year*, June 17, 2019.

[33] Merrill Corporation, *GDPR Burdens Hinder M&A Transactions in the EMEA Region, According to Merrill Corporation Survey* (November 13, 2018), https://www.merrillcorp.com/us/en/company/news/press-releases/gdpr-burdens-hinder-m-a-transactions-in-the-emea-region.html.

[34] Jian Jia, Ginger Zhe Jin, and Liad Wagman, *The Short-Run Effects of GDPR on Technology Venture Investment* (May 31, 2019), http://dx.doi.org/10.2139/ssrn.3278912. For a different view see Lorien Sabatino and Geza Sapi, *Online Privacy and Market Structure: Theory and Evidence*, Dusseldorf Institute for Competition Economics, Discussion Paper 308 (February 2019)(Their model assumes that those with superior technology need less data to achieve similar results, and thus can outcompete their rivals on offering privacy policies. Once all competitors are required to introduce privacy protection, the superior firms lose their comparative advantage. However, the model disregards the different costs imposed on large and small firms in implementing the GDPR, as explored in this article. It also disregards the effects on data-based innovation of the latter's reduced ability to enjoy economies of scale and scope).

analyzes the effects of the GDPR on competition and innovation under diverse conditions, such as the height of economies of scale and scope in data analysis in the specific market, and the ease of collecting similar data or data that is capable of providing relatively similar information. As shown, under some market conditions the GDPR has unintended and so far unrecognized effects on competition, efficiency, innovation, and the resultant welfare. Chapter VI concludes and suggests some means of reducing such effects, while still protecting the vital goal of privacy.

II. SETTING THE STAGE

A. Three Assumptions Regarding Data Markets

Before we delve into the analysis of the GDPR, let us set the stage by briefly reviewing some relevant characteristics of data markets. We then emphasize three fundamental assumptions on which the analysis below is based.

Data markets consist of three main links along the data value chain: collection, processing, and use of data-generated information and knowledge.[35] Collection relates to the extraction of the data and its datafication, namely the recording, aggregation, and organization of information into a form that can be used for data mining, including its transfer to servers.[36] Processing relates to optimizing, cleaning, parsing or combining different datasets, in order to organize the data for future extractions and to find correlations. It can transform the raw data into *information*, defined as data-in-context. Processing may also involve data analytics, thereby creating *knowledge*. Use is defined as employing data-based information or knowledge for prediction and decision-making in relevant markets. It can lead to improved or innovative processes, products, services, and predictions. This value chain also has a dynamic internal reciprocal dimension, in which by a process of machine learning, information regarding the success of the algorithm's past predictions may be used to "teach" the algorithm so that it can make better predictions in the future, thereby creating a feedback loop.[37]

The analysis below is based on three assumptions with regard to data markets, which are based on our current understanding of the information economy. The first relates to the importance of data.[38] As numerous studies have emphasized, with data "rapidly becoming the lifeblood of the global economy,"[39] the efficiency of its use

---

[35] Other links may involve storage, sharing, or deletion/destruction.

[36] See Katherine J. Strandburg, *Monitoring, Datafication, and Consent: Legal Approaches to Privacy in the Big Data Context*, in Julia Lane et al., eds, PRIVACY, BIG DATA, AND THE PUBLIC GOOD: FRAMEWORKS FOR ENGAGEMENT, 10–12 (Cambridge 2014) (discussing the three basic modes of data acquisition); Helen Nissenbaum, *Deregulating Collection: Must Privacy Give Way to Use Regulation*?, *in* DIGITAL MEDIA AND DEMOCRATIC FUTURES, 8–9 (Michael X. Delli Carpini, ed., 2019), archived at http://perma.cc/HP4A-T2G2 (data is "constructed or created from the signals of countless technical devices and systems").

[37] Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 89–95 (2014).

[38] The next two paragraphs are largely based on Michal S. Gal and Daniel Rubinfeld, *Data Standardization*, 94 NYU LAW REV.101 (2019).

[39] European Political Strategy Ctr., European Comm'n, *Enter the Data Economy: EU Policies for a Thriving Data Ecosystem*, 21 EPSC STRATEGIC NOTES 1 (Jan. 11, 2017), https://ec.europa.eu/epsc/sites/epsc/files/strategic_note_issue_21.pdf.

significantly affects both social and private welfare.[40] Predictions based on patterns and correlations identified in data affect numerous aspects of our lives, including health, education, transportation, and sustainability.[41] For some purposes, the collection and use of personal data may be essential.

The second assumption is that while data needs differ from one firm to another and from one industry to another, for many applications the quality of knowledge that can be extracted from data is correlated with the data's *volume* (the amount of data used in the analysis), its *variety* (the diversity of its sources), its *veracity* (accuracy and reliability), and its *velocity* (freshness).[42] This is for three reasons.[43] First, data analysis is often characterized by economies of scale and scope, at least up to a point.[44] This implies that until such economies are reached, the more data is available and the more varied the data, the better the knowledge that can be mined from it. As Mayer-Schönberger and Padova observe, "the value of data can be greatly enhanced...by combining it with other data sources. It is like a single puzzle piece that taken by itself offers little value, but when combined with others to complete an image is turned into something precious."[45] Second, the volume, variety, veracity, and velocity of the data may affect the quality of the algorithm used for its analysis, due to the algorithm's feedback loop, with the algorithm evolving from learning based on an analysis of past predictions.[46] Accordingly, the better the data, the better the algorithm and the better its predictions. Finally, the qualities of a dataset can also create positive externalities with respect to other datasets. This is because of "transfer learning": an algorithm can "learn" from a high-value dataset to perform tasks that can then be performed on other datasets.[47]

---

[40] *See generally* OECD, *supra* note 40 (describing how data now drives all aspects of innovation in the economy and society).

[41] *See, e.g.*, COUNCIL OF ECON. ADVISORS, EXEC. OFFICE OF THE PRESIDENT, BIG DATA AND DIFFERENTIAL PRICING (2015), https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/docs/Big_Data_Report_Nonembargo_v2.pdf (examining the ability of companies to charge different prices to different consumers based on predictions gathered from big data); Federal Trade Commission, BIG DATA: A TOOL FOR INCLUSION OR EXCLUSION? (2016) available at https://www.ftc.gov/system/_les/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf. (increasing some benefits to otherwise underserved populations).

[42] See, *e.g.*, MAURICE E. STUCKE AND ALAN P. GRUNES, BIG DATA AND COMPETITION POLICY (2016).

[43] Gal and Rubinfeld, *supra* note 38.

[44] STUCKE AND GRUNES, *supra* note 42, at 352–55. A debate ensues on the extent of such economies of scope in different markets. With regard to search data see, e.g., Andres V. Lerner, *The Role of "Big Data" in Online Platform Competition* 4–5 (2014), http://ssrn.com/abstract=2482780 (where data is not time sensitive, incremental data beyond a certain volume yields diminishing marginal returns to scale); Lesley Chiou and Catherine E. Tucker, *Search Engines and Data Retention: Implications for Privacy and Antitrust* MIT Sloan Research Paper No. 5094-14 (May 27, 2014), https://ssrn.com/abstract=2441333.

[45] Viktor Mayer-Schönberger & Yann Padova, *Regime Change? Enabling Big Data Through Europe's New Data Protection Regulation*, 17 COLUM. SCI. & TECH. L. REV. 315, 320 (2016).

[46] *See* STUCKE & GRUNES, *supra* note 42, at 170; Maryam Farboodi *et.al., Big Data and Firm Dynamics*, the National Bureau of Economic Research Working Paper No. 25515 (Jan. 2019)(Under data feedback loops, more data leads to increased productivity, enabling firms to sell more and accumulate more data, thus further increasing productivity and creating an indirect network effect).

[47]See, e.g. Lilyana Mihalkova *et al*., *Mapping and Revising Markov Logic Networks for Transfer Learning*, *in* PROCEEDINGS OF THE 22ND CONFERENCE ON ARTIFICIAL INTELLIGENCE (AAAI-07) 608, 608 (2007).

The third assumption is that in many market settings data sharing plays an important role in realizing potential data-based benefits.[48] This is because much data is collected in a system that is largely modular and distributed.[49] For example, as of this writing, thirty billion Internet of Things devices, controlled by numerous market players, are hooked to the internet, collecting and using data.[50] While, once collected, data is non-rivalrous, in many markets entry barriers exist to data collection.[51] In such a system, data sharing may enable more entities to use the data and create data synergies. Furthermore, barriers to data sharing could result in the balkanization of data within particular sectors or firms, thereby not only impeding innovation within markets, but also reducing spillovers to the improvement of analytical tools.[52] Accordingly, broadening and improving the use of data through data sharing is likely to increase the competitive advantages of firms and nations. Data subjects can enjoy some of these benefits through better products as well as online transactions and interactions based on better matching. At the same time, the sharing of personal data can be welfare-reducing due, inter alia, to price increases, price discrimination, and intangible harms such as psychological discomfort and harm to freedom of speech.[53] Sharing of personal data therefore creates complex, often ambiguous, tradeoffs that require a careful and conscious balancing between the competing considerations.[54] Yet at least in some situations, data sharing has significant potential for increasing both private and public welfare.

B. Five Potential Business Models for Obtaining Relevant Data

When data constitutes an important input in a firm's operations, it must choose its strategy for collecting and processing such data. Based on interviews with market participants, we identified five main strategies employed by market players for amassing relevant data, as follows:

1. Organic data collection ("first-party data"): the firm gathers the data directly from the Data Subject;
2. Merging with an entity and using its data in your own operations;

---

[48] See, e.g., Gal and Rubinfeld, *supra* note 38; De Streel and Thomas Tombal, "The Fifty Shades of Data Sharing and the Law' (2019, on file with author); VIKTOR MAYER-SCHONBERGER AND THOMAS RAMGE, REINVENTING CAPITALISM IN THE AGE OF BIG Data (Basic Books, 2018).

[49] *See* GREG ALLEN & TANIEL CHAN, BELFER CTR., HARVARD KENNEDY SCH., ARTIFICIAL INTELLIGENCE AND NATIONAL SECURITY 27 (2017), https://www.belfercenter.org/sites/default/files/files/publication/AI%20NatSec%20-%20final.pdf.

[50] *Internet of Things (IoT) Connected Devices Installed Base Worldwide from 2015 to 2025 (in Billions)*, STATISTA, https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide (last visited Mar. 20, 2019). Some firms enjoy comparative advantages in data collection. *See infra* Section x. The term Internet of Things relates to the digitization of the physical world through the creation of a network of devices (e.g., cars and refrigerators) that contain electronics, software, sensors, actuators, and connectivity which allows them to connect, interact, and exchange data.

[51] Daniel Rubinfeld and Michal S. Gal, *Entry Barriers to Big Data*, 59(2) ARIZONA L. REV. (2017).

[52] Iain M. Cockburn *et al., The Impact of Artificial Intelligence on Innovation*, *in* THE ECONOMICS OF ARTIFICIAL INTELLIGENCE: AN AGENDA 115, 125–28, 139–43 (Ajay K. Agrawal et al. eds., 2019).

[53] Acquisiti et al., *supra* note 13.

[54] *Ibid*.

3. Buying/receiving the data from an external supplier ("third-party data"). This option also includes the sharing of data through the use of an application programming interface (API). An API is an interface or communication protocol between a client and a server such that the server will initiate a defined action, including providing data, in response to a recognized request by the client for data in a specific format. APIs are the mechanism by which, for example, firms can access data held by Facebook, for specific purposes such as authentication, once a data subject gives such firms permission to request such details;
4. Becoming part of a joint venture in which firms pool their data (or data-based knowledge) for specific predefined purposes;[55]
5. Buying/receiving data-based knowledge (rather than data), or aggregated data, from an external provider (knowledge broker). For example, Google provides a service that uses its database to answer queries without exposing the data that served as the basis for the answers. Similarly, market insight services provide reports about general market trends. This option is not subject to the GDPR.

Options 1 and 5 generally do not allow for the realization of inter-firm data synergies. Furthermore, option 5 is an outlier, since it does not enable the firm to use the data internally, and therefore its use is limited to those instances where data-based knowledge will suffice. We will therefore relate mainly to options 1 through 4.

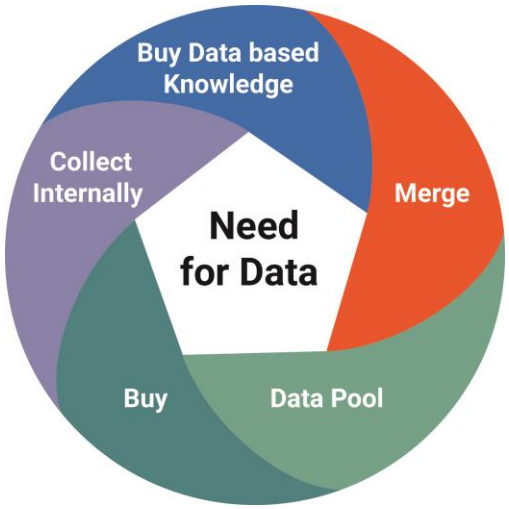

Figure 1: Five options for obtaining necessary data or data-based knowledge

[55] Firms may use the mechanism of joint controllers in accordance with Article 26 if the GDPR. In which case, both Controllers will be jointly and separately responsible and liable for any data processing activity and compliance with the regulation vis-à-vis the data subject.

Firms choose from these options based on their relative cost-effectiveness, as well as their feasibility and scalability, which, in turn, are affected by a combination of technological, financial, strategic, and legal barriers. Technological barriers are those factors which impede the collection or sharing of data—for example, barriers to interoperability between databases which were organized by different entities in accordance with different internal logics. Some such barriers may be overcome, but at a cost—for example, the costs accruing from short-term system shutdowns necessary to enable interoperability, or the costs of data lost due to interoperability limitations.[56] Other technological barriers may be prohibitive, such as the inability to directly collect historical data ex post. Financial barriers are those factors which prevent data from being amassed in a cost-effective way. For example, internal data collection is financially feasible only when its benefits outweigh the costs of putting in place and operating such a system. Strategic barriers are those erected by data owners in order to retain their market power. For example, data controllers may set highly restrictive terms, or may be reluctant to share their data at all, in order to preserve a comparative advantage. Finally, legal barriers are those imposed by laws and regulations relating to the collection, processing, and use of data, the GDPR being a prime example. Legal barriers may affect the other types of barriers. For example, as elaborated below, the requirement that data be transferred at the request of the data subject "where technically feasible"[57] may affect the incentives of firms to adopt technical standards for data portability and interoperability. In the analysis below, we treat non-legal barriers as a given, and focus on how legal barriers affect the choice between the five business models described above. As will be shown, in some cases the legal data regime has a decisive effect on firms' choices, given the regulatory power of the state in shaping potential market interactions.

III. Shaping Choices: The Potential Effects of Legal Limitations on the Five Business Models

A. Legal Obligations Regarding Data Collection, Processing, and Use

The GDPR imposes obligations regarding the collection, processing, storage, and use (including sharing) of data. Below we analyze the main legal limitations that affect firms' choices between the five business models for collecting the data necessary for their operations—choices which, in turn, shape market structures.[58] While the focus is on legal barriers, the analysis will also relate to the technological, financial, and strategic barriers which are affected by the GDPR, where relevant.

As we show, the legal data regime affects the choice between the business models for collecting data. The intuition is straightforward: where data is necessary for a firm's activities, it will seek the most efficient and least costly way to obtain it. For instance, when sharing of data is not allowed or is too costly, firms will have stronger incentives to engage in internal collection, and vice versa. The relative costs of meeting their legal obligations under different business models may therefore lead firms

[56] Gal and Rubinfeld, *supra* note 38.
[57] Article 20 of the GDPR.
[58] The GDPR is still relatively new. Accordingly, our analysis is based on knowledge acquired so far.

to make choices they would otherwise not have made. This is, in fact, a modern example of the Coasean tradeoff between internal and external procurement of products and services.[59]

Let us explore this proposition in greater depth. The stronger the contractual,[60] legal,[61] and reputational[62] sanctions for non-compliance with the GDPR, the more important it becomes for firms which need data for their operations to adopt business models that ensure compliance. In the analysis below we assume that the extent of such sanctions is sufficient to incentivize most firms to attempt to comply. Indeed, beyond the increased fines imposed by the GDPR, we assume that firms today are highly conscious of the reputational damage likely to follow any carelessness with data—a product of heightened recognition of the potential harms resulting from security breaches, coupled with assimilation of the idea that data subjects own their personal data, which together have significantly increased the (theoretical) value data subjects attach to data protection. Furthermore, as elaborated above, non-compliance creates a risk of virality, where non-compliant data can pollute a whole dataset, the algorithms that are trained on it, and the supply chain which relies on it.

To recognize how the GDPR affects firms' choice between business models for data collection, we must identify the different costs, risks, limitations, and benefits it imposes on the different models. Some legal obligations are relevant only to some business models and may therefore create comparative hurdles to adopting them. Other legal obligations apply to all business models. Yet, as shown below, even general obligations may impose dissimilar compliance costs under different models. While it is impossible to canvass in this short article all the relevant legal obligations, we relate to five examples which we regard as having the strongest influence on the choices firms make. These legal obligations apply to more methods for data collection or sharing than in the pre-GDPR era.[63]

1. Ensuring the Lawfulness of Data Processing

---

[59] Ronald Coase, *The Nature of the Firm*, 4 ECONOMICA 386 (1937)(Firms internalize production actions so long as the costs of internal management and production are lower than the transaction costs of market relationships).

[60] For instance, Microsoft's Supplier Security and Privacy Assurance program applies to suppliers that process Microsoft data on its behalf. Payment for services is conditional, *inter alia*, on compliance with data management requirements under the GDPR. Microsoft, Supplier Security and Privacy Assurance, https://www.microsoft.com/en-us/procurement/sspa?activetab=pivot%3aprimaryr3.

[61] A known yet unique example involves a €50 million fine levied on Google by the French data regulator for its "lack of transparency, inadequate information and lack of valid consent regarding ads personalization". The CNIL's Restricted Committee Imposes a Financial Penalty of 50 Million Euros Against Google LLC, 21 January 2019, https://www.cnil.fr/en/cnils-restricted-committee-imposes-financial-penalty-50-million-euros-against-google-llc

[62] As seen, for example, in the anger expressed by Instagram users when Facebook announced changes in the site's privacy policy after acquiring Instagram in 2012. See, e.g., Amy A. Hinkler, *Privacy in the Age of Social Media Mergers and Acquisitions* (2013), https://pdfs.semanticscholar.org/2d3b/014ea02cd0e22305316e87fe8db0dd76b663.pdf.

[63] To illustrate, the GDPR, in conjunction with the ePrivacy Directive, impose limitations on some technological methods of data collection that were once widely used, and that were largely assumed to fall outside the scope of "personal information" as previously defined. These include, inter alia, the use of anonymized web identifiers such as cookies, tags and pixels, when they can lead to identification of a natural person. Recital 30 of the GDPR; Recital 20 and Article B(25) of the ePrivacy Directive 2002, as amended in 2009.

All data processing must meet four fundamental requirements: data must be processed lawfully, fairly and transparently (lawful basis); must be collected for specified, explicit and legitimate purposes only (purpose limitation); must be limited to data required for the entity's defined purposes (data minimization); and must be kept accurate and up-to-date.[64] Lawfulness of data processing is based on six alternative legal bases.[65] The most relevant for our analysis involve a data subject's consent for processing his personal data, and a data controller's "legitimate interest" in processing the data.

Consent is defined as "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her."[66] The requirement for consent is dynamic, in the sense that the more data is collected, or the more sensitive it is to identification or risk of a breach, the stronger, more informed and active consent must be.[67] Observe that consent must relate both to the use being made of the data, and to the entity making such use. The assumption here is that a data subject might be willing to accept a certain use of her data by one specific entity, but not by another. Accordingly, a general consent to "use data for all appropriate purposes," or to "share data with others," would not meet the necessary conditions.[68]

Consent is generally easiest to obtain in the internal collection model, for two main reasons. First, use by the entity which collects the data is more direct and intuitive.[69] Second, consent for use by a certain entity may apply to all its internal units and divisions. To illustrate, if a data subject consents to the use of her data by Google for specified purposes, there is no need to obtain new consent for each of Google's internal units which makes such use of the data, or even for a new internal unit. In economic terms this implies that economies of scale and scope exist in obtaining user consent. This is a major advantage, since obtaining each data subject's explicit and informed consent for all the specific uses of the data pertaining to her is costly. This also implies that large and diversified data controllers, which combine the consent requirements for all their data uses, enjoy an advantage. Google, for example, asks users to actively click a "sign up" button under the notification that such action means consent to the firm's data policy and cookie policy, and pertains to all its uses of the

---

[64] Article 5 of the GDPR.
[65] Article 6 of the GDPR.
[66] Article 4(11) of the GDPR. Additional conditions are specified by Article 7 and recitals 42 and 43. The Article 29 Working Party adopted Guidelines on Consent under Regulation 2016/679 (April 16, 2018). For obstacles to gathering data created by opt-in mechanisms see, e.g., Fred H. Cate & Michael E. Staten, *Protecting Privacy in the New Millennium: The Fallacy of "Opt-In"*, available at http://bit.ly/2lvZ9uz.
[67] *Ibid*.
[68] Vikas Kathuria and Jure Globocnik, *Exclusionary Conduct in Data-Driven Markets: Limitations of Data Sharing Remedy*, *in* EU COMPETITION LAW REMEDIES IN DATA ECONOMY (Marco Botta ed., 2019). See also Article 29, Working Party, Guidelines on Transparency under Regulation 2016/679 (November 29, 2017), p. 9. The requirement for transparency exists independently of the requirement to ensure that there is an appropriate legal basis for data processing.
[69] Here we relate to divisions within one entity. In case of common ownership of different entities ("business groups"), data sharing between entities within the business group encounters higher hurdles. Recital 48 of the GDPR.

user's data.[70] While users can opt out at any time from all or from select uses of their data, this requires an action on their part. Furthermore, if the requesting firm enjoys a reputation for protecting its users' privacy, or users assume the firm already possesses much information about them, consent may be more easily granted for a wider range of purposes, creating a positive externality on new (but consented-for) uses. Moreover, multi-product or multi-service firms have many more entryways through which they can obtain the consent of data subjects to use their data. Indeed, Campbell, Goldfarb and Tucker demonstrate that consumers are more likely to grant consent to large networks with a broad scope rather than to less-established firms.[71] Observe, however, that strong market power adds another layer of complexity to obtaining consent. Consent requires that data subjects have a genuine and free choice as to whether or not their data is to be processed for a particular purpose. At the same time, as Graef observes, when the data controller has significant market power, the consent provided by a data subject might not fully meet those requirements.[72]

Sharing data with an external entity often encounters even higher hurdles in obtaining user consent. It requires the data subject to give her informed consent to the specific use of the data by the external user. In some cases, where such sharing is known, the initial request for consent may already include consent to the use of the data by a specific third party, or at least, categories of third-parties with which the data is being shared. Google, for example, requires that publishers that rely on its services, such as The London Times, request its users to consent to third party tracking by Google, to be used for its own analytical purposes, when they consent to using the publisher's services.[73]

Two owners of databases merge. Each obtained the consent of their data subjects to their use of the data. Are they automatically allowed to merge their datasets, which were created prior to the merger? The answer is negative, unless both merging entities obtained the consent of their data subjects to the use of their data by the merged entity for its specified purposes, or unless they can rely on an alternative basis for lawfulness. Indeed, in the Facebook–WhatsApp case, the attempt to merge data held by the two entities raised breach-of-privacy concerns.[74] A similar obstacle applies in the case of a data-sharing joint venture.

To overcome such obstacles, many privacy policies and consent mechanisms include a condition that personal data may be shared in case of a merger or acquisition, or that the user's consent also applies to sharing the data with a third party for specified purposes. Yet this may not be sufficient where the consent was not sufficiently specific with regard either to the additional use or to the identity of the new user.

---

[70] For some purposes or sensitive data categories, additional consent requirements are imposed. For B2B services, the consent mechanism may be different, and processing of data is often based on a contract.

[71] Campbell *et al.*, *supra* note 7. For a suggestion that consumers are more likely to give consent to large firms see A. M. McDonald and L. F. Cranor, *The Cost of Reading Privacy Policies*, 4(3) JOURNAL OF LAW AND POLICY FOR THE INFORMATION SOCIETY (2008).

[72] INGE GRAEF, EU COMPETITION LAW, DATA PROTECTION AND ONLINE PLATFORMS: DATA AS ESSENTIAL FACILITY (2016).

[73] Testimony of Alastair Mactaggart Chair, Californians for Consumer Privacy, Senate hearings, *supra* note 11.

[74] Irish Data Protection Commission Statement on Proposed Integration of Facebook, WhatsApp and Instagram (January 28, 2019), https://dataprotection.ie/en/news-media/press-releases/data-protection-commission-statement-proposed-integration-facebook.

Should the existing consent not meet these requirements, new consent must be obtained. In addition, where the sharing of data creates a high level of data sensitivity resulting from data synergies, a higher bar to demonstrate consent.[75] Some firms condition the continuation of their service on the user signifying whether and to what extent he consents to such uses of his data. Yet should the service not be in constant use by data subjects, it may take a long time to obtain such consent. Providing new consent might also require an active step by users, something they might be reluctant to do.

Legitimate interests pursued by the data controller or by a third party constitute an alternative basis for lawful data processing.[76] In such cases the controller must perform a tripartite legitimate interests assessment. First, a legitimate interest must be identified. Second, it must be demonstrated that processing is necessary for furthering that interest. Third, the individual's rights must be balanced against the legitimate interest.[77] In the balancing exercise all relevant considerations must be taken into account, including the nature and sensitivity of the data, its uses, and the protective guarantees instituted by the controller to limit the privacy impact. Legitimate interests can be used only when the processing was reasonably expected by the data subject,[78] and its impact on his privacy was minimized.[79]

Legitimate interests are the most flexible basis for lawful use of data under the GDPR.[80] This basis covers a broad range of interests, including situations where data processing is required to prevent fraud, for internal administrative purposes relating to employees and clients, to ensure network and information security, to report possible criminal acts or threats to public security to a competent authority, for direct marketing, and for intra-group data transfers for certain types of data.[81] However, the boundaries of this basis for lawfulness are still unclear. Some suggest that the sale of a customer database is allowed if data subjects are informed about the sale, and allowed to object to the transfer within a reasonable time frame.[82] In one case, the online travel agency TravelBird sold its customer database to its competitor Secret Escapes following the former's bankruptcy. The buyer informed all data subjects about the prospective sale of their personal records and gave them two weeks to object to the transfer. The use of the dataset was limited to the consent given to its original

---

[75] Recital 51 and article 9(1) of the GDPR.
[76] Article 6(1)(f) of the GDPR.
[77] Article 29 Working Party, Opinion 6/2014 on the Notion of Legitimate Interests of the Data Controller Under Article 7 of Directive 95/46/EC (2014), p. 25. While the guidelines relate to the previous Directive, they also provide guidance on the key elements of similar legitimate interest provisions included in the GDPR.
[78] Recital 47 of the GDPR.
[79] Processing that is just "useful" or "convenient" rather than "necessary" will not meet these requirements. P. CAREY, DATA PROTECTION: A PRACTICAL GUIDE TO UK AND EU Law (5th ed., Oxford University Press, 2018), 50.
[80] Opinion of Advocate General Bobek of 19 December 2018 in Case C-40/17 (Fashion ID), ECLI:EU:C:2018:1039, para. 122 (the notion of legitimate interests "appears to be rather elastic and open-ended"); Article 29 Working Party, *supra* note 77, p. 46.
[81] Recitals 47–49 of the GDPR; Article 29 Working Party, supra note 77; Article 29 Data Protection Working Party, Opinion 01/2017 on the Proposed Regulation for the ePrivacy Regulation (2002/58/EC), adopted 4 April 2017, p.7 ("Open-ended exceptions along the lines of Article 6 GDPR, and in particular Art. 6(f) GDPR (legitimate interest ground), should be avoided.").
[82] I. Chu Chao, *How to sell customer data under the GDPR?* Linkedin, https://www.linkedin.com/pulse/how-sell-customer-data-under-gdpr-i-chu-chao

owner, and subject to that owner's privacy policy.[83] Whether such interpretations are correct is a question for the European Data Protection Board[84] and the courts to answer. At a minimum, it seems to us that data sharing with a third party should be reasonably expected by data subjects.[85]

2. Ensuring Compliance of an External Data Provider with the GDPR

GDPR-imposed compliance responsibilities do not stop at an entity's boundaries. Rather, the GDPR imposes legal obligations on a data receiver to ensure compliance with the GDPR by its external data suppliers, with regard to collection and data processing activities which pertain to the shared data.[86]

Given the importance of vetting costs in our analysis, let us elaborate on what they might include. Some costs are direct—primarily the cost of reviewing the dataset (e.g., what types of personal data it contains) and determining whether the data collection and use complies with GDPR requirements (e.g., how was it collected, what consent was given, what privacy and security policies were put in place by the external data provider). Interestingly, a new market has been created for performing such verifications.[87] In addition, many firms appoint a Data Protection Officer (even in situations where this is not mandated by the GDPR) to ensure ongoing compliance and education. Indirect costs include loss of the ability to use data which is non-GDPR-compliant or is based on consent requirements that cannot be met by the data receiver, or the costs of differentiating personal data from other types of data.

The relative magnitude of these costs differs between different business models. To illustrate, a merger requires one-time vetting, whereas periodic buying from an external data supplier requires ongoing monitoring of compliance. Such costs are not relevant to internal data collection or to the sharing of data-based knowledge. Accordingly, the more difficult or costly it is to ensure compliance of a data supplier, the stronger the firm's incentive to adopt those latter business models.

This requirement also has an indirect effect on market structure. To ensure that the supply chain maintains robust privacy and security practices, compliance is often built into the supply contract.[88] From the point of view of the data supplier, the incentive to do business is limited by the costs of proving compliance, of enabling ongoing monitoring of its activities, or of any sanctions imposed by the data receiver should it be found in breach.

---

[83] Chantal Bakermans, *Selling a customer database: insolvency vs. privacy*, https://penrose.law/en/selling-customer-database/

[84] Under Article 70 of the GDPR, the EDPB is empowered to issue guidelines, recommendations and best practices on how to interpret and apply the GDPR, in order to promote a common understanding of European data protection laws.

[85] Recital 47 GDPR states that "[a]t any rate the existence of a legitimate interest would need careful assessment including whether a data subject can reasonably expect at the time and in the context of the collection of the personal data that processing for that purpose may take place. The interests and fundamental rights of the data subject could in particular override the interest of the data controller where personal data are processed in circumstances where data subjects do not reasonably expect further processing."

[86] Articles 14 and 24 of the GDPR.

[87] Sara H. Jodka, The GDPR and mergers and acquisitions: What corporate buyers and sellers need to know, DICKINSON WRIGHT (April 2018).

[88] See, e.g., Microsoft's SSPA, *supra* note 60. While this example involves data processors, similar contractual sanctions are often applied by data sellers.

3. Ensuring that Shared Access or Shared Data is Used in a GDPR-Compliant Way

In addition to the responsibility imposed on data receivers, the GDPR imposes responsibility on data suppliers to ensure that the former meet the latter's commitments towards the data subject with regard to any data shared. While such duties are not specifically enumerated in the GDPR, they arise from the general principle according to which a data controller is liable towards the data subject to ensure that her data is used only in accordance with her consent, including the right to request, at any time, that the data be erased.[89] While the specifics of this requirement are yet to be determined by courts, it seems to us that at a minimum the data supplier must reasonably assume that the data receiver will comply with the GDPR with regard to any data shared. Sharing thus increases the data supplier's level of risk, as he may not have control over the data receiver. The costs imposed by this requirement may include ongoing monitoring, screening, and auditing of data-related activities performed by the data receiver. The higher the risk of breach, the more stringent such actions need to be. The risk of breach is further increased by the fact that combining the supplier's data with the receiver's data might increase its sensitivity. Yet it might be difficult for the data supplier to determine the extent of this risk, given that he may not be exposed to the properties of the data receiver's dataset. To reduce this risk, many firms engage in due-diligence activities and apply contractual audits and sanctions in case of a breach. Such costs are not relevant to internal collection or to mergers, but are highly relevant to joint ventures and data sale transactions.

Liability goes even further. The Court of Justice recently held that a website operator is jointly liable with web technology providers operating on his site for breaches by the latter of the GDPR, even if the website cannot control what personal data is transmitted to or processed by the web technology provider.[90] Resultantly, the risks of sharing access to data are increased.

4. Data Management Requirements

Once data is collected and stored, the GDPR imposes further requirements on all data controllers. For example, data must be accessed by the minimum number of personnel on a need-to-know and need-to-access basis; and data must be stored and processed for the minimum time and to the minimum degree required.[91] Additional

---

[89] An exception exists where the transfer was initiated by the data subject in accordance with his right to data portability. Article 29 Working Party, *Guidelines on the Right for "Data Portability*" (as last revised and adopted April 5, 2017), p.6: "Data controllers answering data portability requests... are not responsible for the processing handled by the data subject or by another company receiving personal data."

[90] European Court of Justice, June 5, 2018, case C-210/16, ULD Schleswig Holstein v. Wirtschaftsakademie Schleswig-Holstein ("Fan Page"); July 29, 2019, case C-40/17, Fashion ID v. Verbraucherzentrale (joint controller responsibility between a website that used a Facebook Like button and Facebook, with regard to those steps of the data processing that were jointly determined). Empirical studies show that post-GDPR third-party tracking by small and medium firms has been reduced. Batikas *et al.*, *supra* note 10, and also sources cited there.

[91] Articles 25(1) and 25(2) of the GDPR outline the organization's obligations concerning data protection by design and by default. This includes the implementation of appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose are processed. In

constraints and controls, including technical and physical limitations and organizational measures—e.g., non-disclosure agreements, permission management, training and education, documentation, and so forth[92]—are imposed to ensure that data is not misused or overly exposed.

Meeting these requirements may require costly internal (re)organization. One major cost relates to the need to create and monitor a system of data flow to manage the data's accessibility and use. Even within one entity, more than one system of data flow might need to be set in place, for example where the consent of different data subjects relates to different uses, and the data needs to be compartmentalized accordingly. Likewise, different uses of data that do not require access to the entire database create a need for an internal technological architecture that can manage and facilitate permissions for internal access to, transfer of, and use of different components of the database. Such measures are needed to reduce the possibility of data breaches, including breaches that might enable the identification of data subjects through integration of different components of a database.

Legally mandated data management processes exhibit economies of scale and scope, which create a comparative advantage for large data controllers relative to small ones. Costs can be shared through the creation of a joint data pool or a merger, or completely avoided in the case of buying data-based knowledge.

Interestingly, market solutions are constantly being developed to both automate GDPR-compliant data management processes and reduce their costs. For instance, the use of Consent Platform Management (CMP) tools has become commonplace in online markets, even in situations in which the use of such tools is not mandated by the GDPR. CMP tools track and transfer information regarding the consent provided—or withdrawn—by each data subject for any purpose, and enable the data supplier, as well as the data receiver, to verify that the required consent was granted. Such tools also reduce some of the liability risks identified previously.

5. Size-Dependent Legal Obligations

The characteristics of the database affect the imposition of additional constraints on data processing and utilization. The larger the volume, variety, velocity, and veracity of the database, and the richer the knowledge that can be extracted from it, the higher its commercial value. At the same time, the greater the chance that data can be used to extract sensitive or identifying information about individuals, the stricter the requirements imposed by the GDPR. In some cases, the sensitivity of the database, or the scale of the data being processed or monitored, will trigger a requirement to appoint a Data Protection Officer.[93] Likewise, companies are obliged to carry out a Data Protection Impact Assessment any time a proposed data processing activity involves

particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons.

[92] Article 32(1) of the GDPR.

[93] Article 37 of the GDPR lists the cases in which a Data Protection Officer (DPO) must be appointed: (a) the controller is a public authority; (b) the core activities of the controller or the processor consist of processing operations which, by virtue of their nature, their scope, and/or their purposes, require regular and systematic monitoring of data subjects on a large scale; or (c) the core activities of the controller or the processor consist of processing special categories of data on a large scale.

a high risk to privacy or data protection.[94] In addition, the larger the number of entities[95] and individuals that have access to the data, the higher the resultant data security risks, and the stricter the requirement to mitigate such risks, including the adoption of data security measures such as encryption, separation of data, and periodic audits.[96] In short, data sensitivity and the potential for security breach are positively correlated with the extent of required security measures, including anonymization and encryption. Some obligations are also correlated with the size of the firm. For instance, companies with fewer than two hundred and fifty employees are not required to keep records unless their data processing is not incidental or involves sensitive information.[97]

These obligations affect all the business models except the last one, buying data-based knowledge. All the other models assume that the firm wants or needs to increase the size of its dataset. However, while these obligations affect the incentive to create larger datasets, they are generally not prohibitive. Furthermore, putting in place a more sophisticated data management system largely involves one-time costs, the main exception being the appointment of a Data Protection Officer.

But more importantly, the magnitude of these obligations increases only up to the point where the maximal requirements are applied (due to the dataset's size and/or sensitivity). Beyond that point, the marginal costs of meeting legal requirements might fall due to economies of scale and scope. This, in turn, may benefit large firms relative to medium-sized ones which must still comply with such requirements. Put differently, the larger the firm, the lower its per-datum compliance costs, relative to smaller firms which must also comply with similar requirements.

B. Effects on the Choice between Business Models

Let us now analyze the combined effects of the GDPR on the choice between business models. To illustrate such effects, we analyze several stylized case studies in increasing order of complexity. While the extent of the costs imposed by the GDPR might differ from one factual scenario to another, general observations can be made.

Assume that Firm A requires a certain type of data for its operations, which can be lawfully collected by it or by an external firm (Firm E).[98] Absent technological, strategic, and legal barriers to data collection and sharing, the choice of business model will be based on the relative costs of internal versus external data collection.[99] If the costs of internal collection by Firm A are lower than those of Firm E, Firm A will collect the data internally; and vice versa: if Firm E can collect the relevant data more cheaply and efficiently, or if it has already collected the data and has incurred

---

[94] Article 35 of the GDPR; Milda Macenaite, *The 'Riskification' of European Data Protection Law Through a Two-Fold Shift*, 8(3) EUROPEAN J. OF RISK REGULATION 506 (2017).
[95] Including parts of the same business group.
[96] Recital 83 of the GDPR defines the general principles behind security of processing: it must correlate with the assessed risk to privacy based on the processing activities used and the personal data categories being processed.
[97] Article 30 of the GDPR.
[98] In our examples, both firms are data controllers. Of course, Firm A can also employ Firm E as a data processor, to act on its behalf. Also, the choice is not dichotomous, as Firm A might choose to collect some of the necessary data internally and buy other data from external controllers.
[99] This stylized example assumes that the costs of collection also include some costs of processing the data.

the sunk costs involved, Firm A will buy the data from Firm E, thereby saving on data collection costs and limiting duplication in collection.[100]

Let us now add the GDPR. Buying data from Firm E adds several types of costs and obstacles. These include, inter alia, the costs incurred by Firm A for ensuring it has a lawful basis for the use of the data. As noted above, a major obstacle may involve obtaining the consent of data subjects to an additional use of their data.[101] Further costs are incurred by Firm A for vetting the compliance of Firm E with the GDPR in collection and processing actions which pertain to the shared data. Furthermore, acquiring an external dataset may require Firm A to reevaluate its own data life cycle to ensure compliance. In addition, the GDPR imposes costs on Firm E, to ensure that Firm A does not use the shared data in a way which was not agreed upon and which infringes its obligations to its data subjects. These costs will be reflected in the price charged by Firm E for the data. The greater the original data controller's loss of control over the data, and the higher the risk of non-compliance, the higher such costs.

These obstacles and costs, in turn, limit the incentive of Firm A to buy data from Firm E or to create a joint data pool. They might also reduce the number of data controllers competing over the supply of data. That is, if the costs of ensuring or demonstrating GDPR compliance are too high, some firms might choose not to engage in data collection for the purpose of data sharing, thereby reducing the number of competitors in the market. These considerations imply that firm A will have fewer options for buying the data from an external provider, thereby further strengthening incentives for (otherwise inefficient) internal collection. This could also lead to more instances of otherwise inefficient internal data collection, and to the multiplication of data collection efforts and costs.

Another potential option for obtaining the necessary data involves a merger between Firms A and E. Absent legal limitations, in the post-merger situation Firm A could have used Firm E's data internally. Yet, as elaborated above, the merging of legal entities does not automatically imply that the merged entity can use data which was collected by one entity prior to the merger. Rather, the parties must ensure that such use meets the requirements for lawfulness under the GDPR. Furthermore, a merger requires vetting to ensure compliance of the target firm with the GDPR. Mergers that would have otherwise been efficient might thus not be profitable or possible. Our analysis comports with the study noted above, which indicated that more than half of European dealmakers surveyed had worked on a merger that did not progress as a result of concerns associated with the GDPR.[102] This number is alarming, for the reasons elaborated below. While some of the reluctance to share data through mergers might be temporary -until the market adjusts to the new law- many of the costs involved in data sharing are here to stay. The costs of a merger might, however, be lower than the cost of continually buying new data from an external supplier. The

[100] Given data's non-rivalrous nature, Firm A might wish to limit Firm E's ability to share the data with its competitors, especially if such data enables them to enjoy economies of scale and scope. If it cannot do so, its incentives to collect the data internally, to retain its comparative advantage, will be increased.

[101] Of course, Firm A might have encountered similar obstacles if it chose to collect the same data internally. Yet, as noted above, in some situations it might be more difficult to obtain consent for a secondary transfer of the data.

[102] Merrill Corporation, *supra* note 33.

main difference is that a merger involves one-time verification of a basis for lawfulness and GDPR compliance, whereas continual buying requires ongoing monitoring of an external supplier.

The legal data regime also affects considerations such as with whom to merge or from whom to buy the data. All else being equal, Firm A will have stronger incentives to merge with or to buy data from a firm which can be more quickly and inexpensively vetted to ensure GDPR compliance. This, in turn, indirectly affects competition, as it could create a comparative advantage for current data suppliers which have already been vetted. Firm A might also prefer a known (reputable) data supplier over an unknown one, or one larger new data supplier over multiple smaller ones.

The above dynamics increase the costs of sharing data, whether through a simple purchase transaction or through a merger or joint venture, and thereby strengthen incentives for internal collection of data even if an external firm can do so more efficiently or has already invested in data collection. In some (extreme) situations legal compliance costs might also limit internal data collection, leading firms to forego data-based activities altogether. In a survey of data protection professionals, one in five respondents said that full GDPR compliance is impossible.[103] Should the (real or perceived) sanctions for such violations be sufficiently high, internal collection will also be foregone.

The GDPR might also affect the breadth of a firm's activities, products, and services. Limitations on data sharing between different entities could strengthen incentives to adopt a business model under which one company controls all products and services in the relevant data ecosystem, especially where data constitutes an important input and can be collected internally. In line with such benefits, Verizon adopted a strategy of acquiring its data flow supply chain (AOL, Yahoo! and more) to better control and utilize consumer data for targeted advertising and service enhancements, similar to Google's acquisition of DoubleClick and similar assets and technologies involved in its ecosystem's data flow.[104] The greater the obstacles created by the legal data regime to voluntary data sharing between separate entities, the stronger the incentives for such internal expansions.

The GDPR might also generate advantages for larger firms, thereby reducing potential competition.[105] This is for several reasons. First, as noted above, large firms enjoy economies of scale and scope in GDPR-compliant data collection and management.[106] Recognizing this effect, the GDPR imposes more lenient requirements on

---

[103] IAPP and Ernst & Young, *Annual Governance Report 2018* (IAPP and Ernst & Young, 2018), https://iapp.org/resources/article/iapp-ey-annual-governance-report-2018/.

[104] Note, however, that these mergers preceded the GDPR.

[105] Some indications for this effect already exist. See, e.g., Mark Scott, Laurens Cerulus, and Laura Kayali, "Six Months in, Europe's Privacy Revolution Favors Google, Face-book," *Politico*, November 27, 2018, https://www.politico.eu/article/gdpr-facebook-google-privacy-data-6-months-in-europes-privacy-revolution-favors-google-facebook;. Wagner, *supra* note 8; Urban *et al.*, *Measuring the Impact of the GDPR on Data Sharing in Ad Networks*, *in* PROCEEDINGS OF THE 15TH ACM ASIA CONFERENCE ON COMPUTER AND COMMUNICATIONS SECURITY (2020).

[106] Campbell *et al.*, *supra* note 7, p. 48-9: "[R]ather than increasing competition, the nature of transaction costs implied by [opt-in] privacy regulation suggests that privacy regulation may be anti-competitive….[I]n some cases where entry had been profitable without regulation, [some firms] will choose not to enter." Market statistics provide empirical support for this effect with regard to digital advertising, at least in the short run. According to Ghostery, following the introduction of the GDPR, Google's tracking software appeared on slightly more websites, Facebook's on 7% fewer, and the top 50 ad vendors (excluding Google and Facebook) appeared on 20%

small and medium firms. Yet in some situations such lower standards might disadvantage such firms relative to data suppliers which must meet the higher standards. This might be the case when Firm A intends to integrate Firm E's data with its own, creating a large dataset which must comply with the stricter requirements. Firm A might thus prefer to buy data which already complies with the higher standards. Should the added value to Firm A of acquiring such data be significant, a small or medium-sized Firm E might voluntarily choose to meet the higher standards, thereby foregoing the main solution of the GDPR to the competitive disadvantages it creates for small firms relative to large ones.[107] Should the costs of compliance be prohibitive for small firms, the number and even the quality of available data sources might be reduced. It should be noted, however, that the growing development of GDPR-compliance tools, including legal-tech products for automating parts of the compliance process, some of which are open-source, help reduce such costs.

Second, some provisions of the GDPR create uncertainty, which may impose higher costs on smaller players. To illustrate, while the GDPR recognizes a data subject's right to data portability, and mandates that the data be transferred in a structured and commonly used machine-readable format,[108] it does not determine how data should be transferred—for example, which technological standards should be applied to ensure data portability and interoperability. A small player might be concerned that once such standards are set, his costs of data transfer might be high. While this concern affects large players as well, they enjoy two comparative advantages. First, compliance with the standard might be subject to economies of scale, as any costs will be spread over more data. Second, in the absence of governmental facilitation, standards for data operability and interoperability might be set by the large players, which would cater to their own preferences,[109] including raising the compliance costs of their smaller rivals. The absence of standards for organizing a dataset might also increase technological barriers to data interoperability between willing contractual parties.[110] Additional forms of uncertainty which could create a comparative advantage for those with significant resources relates to the boundaries of lawfulness for data use, elaborated above, as well as the exact territorial scope of the GDPR.[111]

Third, uncertainty with regard to the correct interpretation of the GDPR could lead to strategic use of interpretation as an offensive tactic. Following the introduction of the GDPR, some large technology firms limited some of their previous sharing practices. Some commentators argue that these limitations go far beyond what is needed to comply with the GDPR, and thus appear to be used strategically.[112] Geradin and Katsifis, for example, argue that Google's recent limitations on the transfer of

---

fewer, while the smallest companies suffered a 32% drop. Greg Ip, *Beware of the Big Tech Backlash*, WALL STREET JOURNAL, Dec. 19, 2018.

[107] See discussion of the Commission's Impact Assessment, *infra*.

[108] Recital 68 and Article 20 of the GDPR.

[109] Gal and Rubinfeld, *supra* note 38.

[110] *Ibid*.

[111] Batikas *et al.*, *supra* note 10; Oisin Tobin at al., *GDPR: 3 Areas of Ambiguity*, 20(2) PRIVACY & DATA PROTECTION 15 (2019).

[112] For a list of Google policy changes which it justified on the basis of privacy considerations see J. Hercher, *How We Got Here: A Look Back At The Privacy Changes That Reshaped Google*, ADEXCHANGER, 7 October 2019, available at https://adexchanger.com/online-advertising/how-we-got-here-a-look-back-at-the-privacy-changes-that-reshaped-google/

data relating to bids for advertising space could deliver a fatal blow to the market in header bidding (advance bidding on online display ads), since publishers will no longer be able to measure and analyze the incremental value brought by header bidding partners compared to Google-controlled channels. They further claim that, to the extent that Google relies on privacy concerns to justify its policy, the GDPR does not provide such justification.[113]

Fourth, a firm's size, as well as its reputation, may affect the conduct of data subjects. So far, our analysis has largely focused on the incentives of data controllers. But data subjects' willingness to allow firms to collect their data also affects the dynamics of data markets. The GDPR does not directly affect data subjects' choices, except to ensure that they have more information and control over their data. But the increased sensitivity to privacy and human dignity concerns that surrounded its adoption might have such effects. This is because firms which have large datasets might be able to make a more credible commitment to their users that they will follow the GDPR. There are two reasons for this. First, it is rational to assume that those who own large datasets will be monitored more closely by regulators and by their data subjects.[114] Second, firms with high-value datasets will generally have much more to lose if they do not comply with the GDPR than firms with low-value datasets. Similarly, the more important personal data is to one's business model, the larger his potential loss. This is not only because of the size of fines imposed, but, more importantly, because non-compliance could create a backlash among users, at least where they have reasonable alternatives. Users might therefore prefer to engage with firms that have more to lose. This dynamic is similar to that which applies to data security issues, and is strengthened by it. While large technology companies are not immune to hacking, they have strong incentives and hefty funding to erect barriers to hackers and to immediately correct any failures. Accordingly, if the user can choose from several services which are relatively similar, especially if they do not compete on price and/or if non-price aspects are not easily verifiable and quantifiable by the user, the services of large, reputable firms may be preferred.[115] This, in turn, will create barriers to data collection by small or new firms, and further enlarge the datasets of large data controllers. While the GDPR attempts to limit such dynamics by increasing trust that data will be handled by all market participants in ways which protect privacy, as elaborated below. As of yet trust has not increased in the post-GDPR period.[116] In addition, users might also prefer to limit the number of firms to which they grant consent, thereby sticking with a firm to which they have already granted consent for data collection, thereby further strengthening large incumbents.[117]

The analysis so far has implicitly assumed that the data which can potentially be collected by Firm A and Firm E is interchangeable. Let us now focus on a case in which data synergies exist between the datasets of Firm A and Firm E, such that the

---

[113] Damien Geradin and Dimitrios Katsifis, *'Trust me, I'm fair': Analysing Google's Latest Practices in Ad Tech from the Perspective of EU Competition Law* (2019).

[114] The GDPR empowers data subjects to enforce their rights even without court proceedings. Articles 57, 78–80 of the GDPR.

[115] This effect may be tempered, at least with regard to some large technology firms, by surveillance capitalism concerns.

[116] See discussion *infra*.

[117] Roslyn Layton, *10 Problems with the GDPR*, Written Testimony, in Senate Hearings, *supra* note 11.

quality of the information and insights that can be extracted from the combined dataset is significantly greater than that which can be extracted from either alone. Such data synergies can be realized through buying the data, creating a joint venture, or through merger. The legal data regime does not apply different rules to scenarios in which data synergies can be realized and those in which they cannot. Realizing data synergies might, however, involve additional costs and risks relative to simply using an externally created database. For example, as elaborated above, should Firm A combine internal and external datasets which apply different privacy policies, Firm A may need to redesign its internal data flow architecture and its internal policies relating, inter alia, to issues such as who can see what, what data can be used for what purposes, how data is housed and for how long, or what data should be destroyed or separated.[118] Data synergies could also affect the extent of applicable legal requirements. For instance, the newly created combined dataset might include more sensitive data, requiring the firm to meet higher standards for its protection. In extreme cases data which is overly sensitive may be intentionally excluded from data transfers, anonymized or aggregated.[119] The combined dataset might also trigger a requirement to appoint a Data Protection Officer and to conduct a Data Protection Impact Assessment.[120]

Table 1 displays the main costs and benefits of the different business models which involve actual data (rather than data-based information), as affected by the GDPR.

Business model / Costs and benefits[121]	Internal collection	Merger	Buying data	Joint venture
Extending the use of the data	Yes	Yes	Yes	Yes
Compliance costs imposed on data user	Intermediate	Intermediate (one-time)	High (ongoing)	High (ongoing)
Risk of non-compliance by third party that cannot be easily detected	Low	Low to intermediate	High	High
Enabling data synergies	Partial (only if data can be collected)	Yes	Partial (if allowed to sell or share)	Yes
Duplication of direct collection	Yes	No	No	No

Table 1: Main costs and benefits of business models which involve actual data

To summarize the findings of this chapter, the GDPR affects the choices firms make with regard to the source of their data. By imposing high hurdles on data shar-

[118] Darren Abernethy, *Privacy and Data Security in Mergers & Acquisitions*, Linkedin, October 20, 2017, https://www.linkedin.com/pulse/privacy-data-security-mergers-acquisitions-esq-fip-cipp-t-e-a-m
[119] *Ibid*.
[120] See *supra*, footnote 93.
[121] The comparison is based on the assumption that the datasets created are similar, so that monitoring and security costs are similar.

ing, the GDPR strengthens incentives for firms to collect data internally, where feasible, or to buy data-based knowledge if such is sufficient for their needs. It also bestows comparative advantages on large data controllers relative to small ones.

IV. POTENTIAL LEGAL AND TECHNOLOGICAL SOLUTIONS

Before analyzing the welfare effects created by the above dynamics, we explore whether the obligations imposed by the GDPR can be overcome based on legal obligations to share data, or by technological means.

A. Legal Obligations to Share Data

The legal data regime includes obligations for the sharing of data between certain entities, with the aim of reducing access barriers to data in some situations. A prime example is incorporated in the GDPR itself.[122] The right to data portability[123] mandates that data controllers transfer any data they control with respect to a specific data subject, in accordance with the data subject's request that they do so.[124]

In practice, this right is quite limited in its competitive effects. Specifically, it requires that each data subject actively request transfer of his data to another entity. This creates hurdles to the transfer of large quantities of data pertaining to different data subjects, and certainly of whole databases, thereby potentially limiting the ability of the data receiver to enjoy economies of scale and scope in data analysis.

Competition law imposes an obligation to share data in a very limited set of circumstances, in which such sharing is essential for the competitive process. [125] Under the essential facilities doctrine, which forms part of the duty to deal imposed on dominant firms, in exceptional circumstances a dominant undertaking might be mandated to grant access to an input which it controls, and which is indispensable for carrying on a certain business. In *IMS Health*[126] the Court stated that such exceptional circumstances exist where the refusal prevents the emergence of a new product for which there is potential consumer demand, it is unjustified, and it excludes any competition within a secondary market. Later cases, including *Microsoft*, relaxed the requirement for the emergence of a new product.[127]

Significant differences exist between sharing obligations imposed by competition law and those mandated through the GDPR. The values protected are partially different:[128] the GDPR's right to data portability is based mainly on the data subject's ownership of his personal data, while competition law serves to protect and facilitate competition and its outcomes. Also, the initiator of the mandatory transfer differs:

---

[122] A comparable right exists with regard to specific types of data or specific industries, a prime example being the fintech industry.
[123] Article 20 and recital 68 of the GDPR.
[124] Article 20(1) of the GDPR.
[125] For an excellent analysis of this right see Graef, *supra* note 72; Niamh Dunne, *Dispensing with Indispensability* (2019), available at https://ssrn.com/abstract=3476938.
[126] *IMS Health GmbH & Co. OHG v. NDC Health GmbH & Co. KG*, C-418/01, ECLI:EU:C:2004:257.
[127] Case COMP/C-3/37.792 - *Microsoft*, 24 March 2004.
[128] Graef, *supra* note 72.

while under the GDPR the initiator is the data subject, under competition law it is a competition authority or a court.

In rare cases an obligation to share data has been imposed in order to advance competition. A known example involves the decision of the French Autorité de la Concurrence in *GDF Suez*.[129] There, the dominant firm enjoyed a data-based advantage which derived from its previous exclusive access to data on customers obtained within the framework of its former government-created monopoly status. It was ordered to give competing market players access to certain data about its customers (e.g., names, addresses, telephone numbers and consumption profiles), which was deemed essential to enable rivals to compete effectively with it.

Mandatory data sharing through competition law is, however, very limited, for two reasons. First, the conditions for mandatory data sharing are hard to meet. Given that data is often available and non-rivalrous, a competitor might be able to collect similar or comparable data, which could lead to similar predictions. Accordingly, not granting access to data will generally not create a "complete foreclosure of downstream competition" or stifle a "new product" or innovation, as required by the case law.[130] Observe, however, that mandatory data sharing could also be imposed as a remedy, in order to restore competition that has been compromised because of illegal practices by the dominant undertaking.[131] In such cases, the requirements might be more flexible.

Second, competition law regulators have emphasized, time and again, that competition law considerations apply in parallel to data protection considerations, and do not override them. For example, in its *Microsoft/LinkedIn* merger decision, pertaining to the ability of the two firms to combine their datasets, the European Commission clearly stated that any data transfers between the merging entities must also meet the conditions set out in the GDPR.[132] A similar policy is likely to apply to essential facility cases. Given that sharing personally identifiable information—even if it is essential for competition—without a legal basis under the GDPR would impinge the fundamental right to privacy,[133] it is difficult to believe that competition law will take precedence over such rights where a direct clash exists.

At the same time, the fact that one of the bases for data processing under the GDPR involves a legal obligation to process personal data[134] might serve as a tool for reaching a more nuanced and welfare-increasing balance between the two sets of con-

---

[129] Autorité de la Concurrence, Décision 14-MC-02 du 9 Septembre 2014 Relative à une Demande de Mesures Conservatoires Présentée par la Société Direct Energie dans les Secteurs du Gaz et de l'Electricité, paras 169-174, available at http://www.autoritedelaconcurrence.fr /pdf/avis/14mc02.pdf. For a relatively similar Belgian case see Kathuria and Globocnik, *supra* note 68.

[130] Graef, *supra* note 72; Kathuria and Globocnik, *ibid*.

[131] Kathuria and Globocnik, *ibid*.

[132] COMP/M.8124 – Microsoft/LinkedIn, para. 167: "Microsoft is subject to European data protection laws which limit its ability to undertake any treatment of LinkedIn full data...the Commission notes that the newly adopted GDPR... may further limit Microsoft's ability to undertake any treatment of LinkedIn full data by strengthening the existing rights and empowering individuals with more control over their personal data (i.e. easier access to personal data; right to data portability; etc.)." See also COMP/M.8180 – VERIZON/YAHOO, para. 80 et seq. (any combination of the said datasets could only be implemented to the extent allowed by applicable data protection rules); COMP/M.8251 BITE/TELE2/TELIA LIETUVA/JV, para. 85 et seq.

[133] Article 8 of the EU Charter of Fundamental Rights.

[134] Article 6(1)(c) of the GDPR.

siderations, especially with regard to the use of more lenient procedures for data sharing. An interesting example involves the *GDF Suez* case, noted above. GDF Suez was ordered to notify customers about the sharing of their personal data with third parties, and to give them the option of opting out within 30 days. The case was dealt with under the 1995 Directive, under which the requirements for consent were more lenient. While an opt-out mechanism would most likely not be captured under the stricter requirements of the GDPR, in which consent must be explicit and involve an affirmative action,[135] it remains to be seen whether such an opt-out scheme might be legitimate under the legal obligation to process personal data.[136]

Accordingly, both the GDPR's right to data portability and competition law's refusal to deal generally do not significantly affect the choice between business models for obtaining relevant data.

B. Technological Solutions

The choices firms make between the business models are also affected by the technological tools which are available to overcome privacy-based limitations to data sharing. The major tool which can be applied in all five business models, involves data anonymization in ways that make it unreasonable that a third party could de-anonymize the data and identify the data subject or any association with him. In accordance with the European Commission, "personal data that has been rendered irreversibly anonymous in such a way that the individual is no longer identifiable, is not considered to be personal data and thus not governed by the GDPR."[137] Algorithms or other tools that can perform irreversible anonymization are already employed, and in some instances their use is mandated by the GDPR.[138] The market for such algorithms and tools is constantly developing. Some algorithms change the content of the database in such a way that anonymization is ensured, yet any analytics performed on the dataset as a whole are similar to those that would be applied to the original dataset. Others offer tools which make data harder to de-anonymize. Yet complete anonymization is often technically difficult, and sometimes impossible.[139]

This technological solution suffers, however, from several major downsides. First, in many situations data loses much of its value in the absence of personal details about the data subject. For example, the less is known about the preferences of the

---

[135] Kathuria and Globocnik, *supra* note 68.

[136] Graef argues that such measures may fall under the legal obligation exception. Graef, *supra*, note 72, at, 319. For a different view see Kathuria and Globocnik, *ibid*, at 21 and sources cited there, relying, inter alia, on Recital 45 of the GDPR, which requires that the legal obligation be grounded in legislation.

[137] European Commission, *The GDPR: New Opportunities, New Obligations* (Luxembourg: Publications Office of the European Union, 2018), p. 5, https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-sme-obligations_en.pdf

[138] The GDPR requires controllers to use pseudonymization and anonymization when processing personal information. Recital 26 of the GDPR defines anonymized data as "data rendered anonymous in such a way that the data subject is not or no longer identifiable." Data anonymization is a higher bar of security than pseudonymization. Recital 83 of the GDPR mentions encryption as additional security measure which is required to mitigate high risks from data processing activity. Encryption on top of pseudonymization or anonymization carries additional costs in monetary terms and in opportunities to utilize the data.

[139] See, e.g., Article 29 Working Party, Opinion 05/2014 on Anonymisation Techniques, WP 216, 10 April 2014; S. E Fienberg, A. Rinaldo, and X. Yang, *Differential Privacy and the Risk-Utility Tradeoff for Multi-Dimensional Contingency Tables*, in PRIVACY IN STATISTICAL DATABASES (Springer, 2010), pp. 187–199.

data subject, the lower the quality of the match between those preferences and the personalized products directed at that user. This may reduce the firm's profits and potentially even the user's welfare.[140] Second, such solutions limit the possibility of creating data synergies resulting from the direct combination of two or more datasets, rather than combining the meta-knowledge resulting from them. Finally, the acts of anonymizing and categorizing constitute a form of processing of personal data in themselves and therefore must comply with all the relevant requirements.[141] Accordingly, in most situations anonymization solutions do not constitute a viable tool for completely or cost-effectively overcoming GDPR's constraints on data collection and sharing.

V. POTENTIAL EFFECTS ON COMPETITION AND WELFARE

Why should we care if the legal data regime affects business models for data collection and limits data sharing? We suggest several cumulative answers to this question. First and foremost, legal limitations could *prevent the realization of data synergies* between databases controlled by different entities. In a world with many distributed databases, such limitations may significantly weaken the ability to mine knowledge from data.[142] To illustrate, consider medical data on patients' responses to a treatment for a rare disease. Unless data from different controllers are combined into a coherent dataset, doctors will be deprived of means to reach a better understanding of how to treat the disease, for example by learning about unexpected interactions between different compounds. Legal limitations on data sharing could thus prevent the emergence of better data-based products or services, and could reduce firms' ability to develop and fine-tune new algorithms, which might increase productive and dynamic efficiency in a wide array of markets through processes such as transfer learning.[143] These disadvantages are likely to be aggravated to the extent that geographic limitations on data storage erect additional obstacles to the realization of data synergies. It is thus not surprising that barriers to data portability and interoperability between firms have been identified as major impediments to the efficient operation of our data-based economy.[144]

Second, legal limitations on data collection, processing and use could *prevent firms from obtaining the data necessary for their operations.* Assume that Firm A cannot collect all the types of data it needs, or that such collection is not cost-effective. As elaborated by Rubinfeld and Gal, this might be the case where data collection is a side-effect of a costly action (e.g., operating a successful social network), where data is historical (e.g., data about a data subject's past actions), or where data relates to a

---

[140] For an overview of benefits and costs of better matching see, e.g., Acquisiti *et al.*, *supra* note 13.
[141] Graef, *supra* note 72. Article 5 of the GDPR is applicable.
[142] *See, e.g.*, OECD, *supra* note 40, at 29; Gal and Rubinfeld, supra note 38.
[143] Cockburn *et al.*, *supra* note 52 (relating to transfer learning). For an early argument regarding the costs of opt-in privacy mechanisms see Michael E. Staten & Fred H. Cate, *The Impact of Opt-In Privacy Rules on Retail Credit Markets: A Case Study of MBNA*, 52 DUKE LAW JOURNAL 745 (2003).
[144] DIRECTORATE-GEN. OF COMMC'NS NETWORKS, CONTENT & TECH., EUR. COMM'N, STUDY ON EMERGING ISSUES OF DATA OWNERSHIP, INTEROPERABILITY, (RE-)USABILITY AND ACCESS TO DATA, AND LIABILITY 292–93 (2018), https://www.wik.org/fileadmin/Studien/2018/EU_Data_ownership_en.pdf, at 15-16, 88; Oscar Borgogno & Giuseppe Colangelo, *Data Sharing and Interoperability: Fostering Competition Through APIs* (2018)(on file with author).

unique activity which cannot be externally observed (such as data collected on reactions to a certain event which is internal to a social media firm).[145] To engage in its data-based activity, Firm A must therefore obtain the relevant data from Firm E. If obstacles to data sharing are too high, Firm A will not be able to obtain the relevant data, or sufficient data in order to enjoy economies of scale and scope in data analysis. Legal obstacles to internal data processing and use would have a similar effect. [146] In addition, as Layton argues, the need to state at time of data collection the purpose of its use harms the ability to use machine learning techniques to glean from the data unanticipated learnings.[147] The ultimate outcome is further reductions in the availability of data-based products and services.[148]

Some researchers suggest a mitigating dynamic which might enable firms to more easily collect data: by reassuring potential users that their data will be safe, data protection legislation promotes the use of information technology.[149] However, the size of this effect is questionable. A survey administered by researchers at the University of Amsterdam just after the GDPR came into force found that only 1% of 500 respondents had changed their settings and narrowed their consent for the use of their data by any firm.[150] Accordingly, despite the public atmosphere surrounding the adoption of the GDPR, which emphasized the possible harms arising from breaches of privacy, and despite the fact that the legislation made it easy for users to change their consent boundaries by requiring data controllers to actively give each user this option, the new regulation has not necessarily brought about a significant change in users' conduct with regard to their privacy, emphasizing, once again, what some call the privacy paradox.[151] Of course, the GDPR includes additional mandatory privacy-protection mechanisms, but the breadth of consent given by users nonetheless significantly affects what can be done with their data. Furthermore, it is unclear whether the existence of the GDPR provides the reassurance claimed by its proponents: a study conducted by the European Commission in March 2019, more than a year after the

---

[145] Rubinfeld and Gal, *supra* note 51.

[146] Campbell *et al*.. *supra* note 7; Jian, Jin and Wagman, *supra* note 34, p.4 (such obstacles may also arise from uncertainty in the regulation).

[147] Layton, *supra* note 117, p. 6 (pointing to a comprehensive study on the connection between the use of mobile phones and brain cancer, which made use of data which was not specifically collected for this purpose.)

[148] On the connection between privacy and innovation *see, e,g*., Avi Goldfarb and Catherine Tucker, *Privacy and Innovation, in* INNOVATION POLICY AND THE ECONOMY (Josh Lerner and Scott Stern eds., 2012), , Volume 12, p. 65; Silvana Krasteva et al., *The 80/20 Rule: Corporate Support for Innovation by Employees*, 38 INT. J. INDUS. ORG. 32 (2015) (Privacy compliance costs deter innovation and shift some of it into established firms).

[149] Acquisiti *et al*., *supra* note 13. This is part of the logic on which the GDPR is based. The size of this effect may very depending on the types of data involved. *See, e.g.,* I. Adjerid, et al.,*The Impact of Privacy Regulation and Technology Incentives: The case of Health Information Exchanges*," 62(4) MANAGEMENT SCIENCE 1042 (2016)(more privacy of health information increases willingness to provide such data).

[150] Based on discussions with Prof. Balazs Bodo from the University of Amsterdam. Researches have shown that users have a strong tendency to stick with default privacy settings, even when they understand the potential harms involved. See, e.g., Lior J. Strahilevitz & Matthew B. Kugler, *Is Privacy Policy Language Irrelevant to Consumers?,* 45 J. LEG. STUD. 569 (2016); Adam S. Chilton & Omri Ben-Shahar, *Simplification of Privacy Disclosures: An Experimental Test*, 45 J. LEG. STUD. 41 (2016); Jane Bambauer et al., *A Bad Education*, 2017 IL. L. REV. 109 (2017).

[151] The privacy paradox is the discrepancy between individuals' declared attitudes (concerns about threats to their online privacy) and their actual behavior (freely disclosing personal data online). See, e.g., Susanne Barth and Menno D. T. Jong, *The Privacy Paradox – Investigating Discrepancies Between Expressed Privacy Concerns And Actual Online Behavior – A Systematic Literature Review*, 34(7) TELEMATICS AND INFORMATICS 1038 (2017).

GDPR came into force, found that 81% of Europeans who reported providing data online still said they felt they had no control or only partial control over their data.[152] Accordingly, the effects of this mitigating factor do not currently seem sufficiently strong to increase competition in the market. It remains to be seen whether this will still be the case once users are better educated regarding their ability to control their data, and once significant enforcement actions are undertaken and made public.

Third, limitations on data sharing may reduce competition and *lead to more concentrated market structures*.[153] Two dynamics are at play here. First, as elaborated above, the GDPR creates some comparative advantages for large firms in meeting GDPR requirements.[154] Second, limitations on data collection and sharing reduce the potential for the emergence of a competitive and distributed data collection ecosystem. To illustrate this point, compare two scenarios. Under the first, data sharing is relatively easy. This increases the incentives of firms to collect data. As a result, the market for data may become more competitive. Furthermore, the increased ability of firms to integrate different datasets reduces the need to rely on one source for data, either internal or external, thereby lowering the price for data.[155] Under the second scenario, high obstacles to data sharing reduce the ability and incentives of some data controllers to enter or operate in the market. A potential indication of this effect can be found in the fact that some foreign firms have exited European markets in order to avoid being subject to the GDPR,[156] or that the number of small and medium sized competitors in some digital markets was significantly reduced,[157] thereby reducing potential competition. While some firms may return once the dust has settled on the implications of GDPR compliance, some may not.

The above dynamics, in turn, could strengthen one of the main concerns raised in digital markets: the fact that some entities enjoy significant and durable market power based largely on their control of vast amounts of data.[158] The comparative advantages enjoyed by such firms are partly based on economies of scale and scope in data collection and analysis, and on network effects.[159] Other firms, which lack such a variety and volume of data sources, may find it difficult to match these capabilities, especially where first-mover advantages and switching costs are high.[160] This difficulty may be overcome if the competitor could combine data collected by numerous sources. And

---

[152] European Commission, *Special Eurobarometer 487a, The General Data Protection Regulation* (June 2019), http://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/survey/getsurveydetail/instruments/special/surveyky/2222. It also found that "at a country level there is no consistent relationship between awareness of GDPR and the level of control respondents feel they have over the personal information they post online."

[153] This paragraph and the next largely build on Gal and Rubinfeld, *supra* note 38.

[154] For an empirical study see Batikas et al., *supra* note 10.

[155] For the tradeoffs between internal and external relationships in interconnections *see* Ronald H. Coase, *The Nature of the Firm*, 4 ECONOMICA 386, 394–95 (1937).

[156] Alec Stapp, *GDPR After One Year: Costs and Unintended Consequences*, TRUTH ON THE MARKET, 24 May 2019 https://truthonthemarket.com/2019/05/24/gdpr-after-one-year-costs-and-unintended-consequences/; Layton, *supra* note 117, p. 3-6 ("No longer visible in the EU are more than 1,000 American news and media outlets, in addition to many sites for ecommerce, games, information technology, and other services." This also indirectly affects free speech and democracy).

[157] Layton, *ibid.*

[158] *See, e.g.*, OECD, *supra* note 40, at 7; John M Newman, *Antitrust in Digital Markets* (June 22, 2018), https://ssrn.com/astract=3201004.

[159] STUCKE & GRUNES, *supra* note 42; Howard A. Shelanski, *Information, Innovation, and Competition Policy for the Internet*, 161 U. PA. L. REV. 1663, 1679 (2013).

[160] *Id.*

since data is non-rivalrous and often easily replicable, data controllers could potentially share their data with many users, thereby strengthening competition even further. Yet if firms encounter high hurdles to combining data collected by external sources, they may not be able to enjoy economies of scale to the same extent as established rivals.[161] This, in turn, could entrench market power and raise the risk of monopolization.[162]

Finally, European firms' *international competitiveness* might be negatively affected.[163] Leading in artificial intelligence-based products and services requires more and higher-quality data from which algorithms can learn. Accordingly, even when data collected in one country is not relevant to consumers in another, limitations on data sharing and synergies could reduce the ability of domestic firms to create better algorithms, and could impede transfer learning and its application to foreign data. The concern for competitiveness is strengthened by the fact that some governments have started to invest in creating ecosystems for data-driven advantages. China, in particular, has begun seeking to create such advantages for its domestic firms by, inter alia, motivating the creation of huge, comprehensive databases in areas where big data is considered to be of utmost importance (e.g., medical devices, autonomous cars, smart cities).[164] Moreover, stricter limitations on data sharing with foreign entities and geographic limitations on data storage[165] affect the incentives of foreign entities to engage in data sharing with European firms, thereby potentially creating opportunity costs resulting from foregone business opportunities. Opportunity costs could also arise from the employment of system engineers to ensure GDPR compliance.

Accordingly, while privacy protection is of utmost importance for ensuring the well-being of European citizens, more thought should be given to its side effects and to designing welfare-enhancing balances between the competing considerations. In this regard, we emphasize once again that while the costs of putting a GDPR compliance system in place are short-term, other limitations imposed by the GDPR are long-term. Accordingly, while in the future we may see more data-sharing deals and more entry than in the current adjustment phase, most of the dynamics identified in this article reflect ongoing compliance with GDPR requirements, and so are here to stay.

Interestingly, the European Commission's own Impact Assessment on the GDPR, published in early 2012, concluded that while the GDPR would impose limited costs (mainly for the employment of a data protection officer, where relevant, and the introduction of a general obligation to demonstrate compliance, which was estimated to

---

[161] *Id.* at 166.

[162] For the risks of entrenchment of market power resulting from opt-in privacy consent mechanisms see Campbell et al., *supra* note 7, p. 68 ("privacy regulation can shield a large, general incumbent from potential competition because regulation raises the threshold quality and scope for profitable entry by a challenger…. This is more likely for relatively strong incumbents: the stronger the incumbent, the better the marginal entrant must be").

[163] For a similar argument with regard to data standardization see Gal and Rubinfeld, *supra* note 38.

[164] Cheng Yu and Ma Si, *Industrial Internet to Boost Smart Manufacturing*, CHINA DAILY (Dec. 1, 2017, 7:35 AM), http://www.chinadaily.com.cn/business/tech/2017-12/01/content_35148829.htm.

[165] Recital 101 and Article 45 of the GDPR. For instance, geographic limitations on data storage may erect barriers to international transfers of data which might be needed in order to create a joint dataset (e.g., on patients' reactions to a specific treatment) that includes data on data subjects from EU and third countries/foreign firms which have not received the required EU approval. In some instances such barriers can be (partially) overcome by storing the relevant data in the respective countries and querying the different datasets separately.

impose "some additional" administrative burdens), it would also strengthen competition overall, and would support the competitiveness of EU firms in numerous ways.[166] These were expected to include lowering the costs of compliance with privacy requirements by replacing then-existing EU data protection laws with a clear, simple regime and consistent and effective enforcement, and facilitating business by increasing consumer trust.[167] With regard to its impact on innovation, it was concluded that increased consumer trust under the GDPR would also increase the uptake of new technologies, with the result that European industries could become world leaders in privacy-enhancing technologies or privacy-by-design solutions.[168] Addressing industry concerns regarding the administrative and financial costs of implementing some of the proposed changes, the Impact Assessment emphasized that some of the GDPR's legal obligations would not be imposed on small firms.[169]

The Commission's Impact Assessment disregards important factors such as the effects of the GDPR on the ability of firms to enjoy data-based advantages by way of data synergies and economies of scale and scope in data analysis, which may significantly affect competition and innovation. One of the reasons for this truncated assessment was that it focused on comparing the GDPR to the previous privacy directive, adopted in 1995. Accordingly, the GDPR's overall impact on internal competition and international competitiveness in a digital environment was not evaluated. This is unfortunate, as the analysis disregards important aspects of the GDPR's effects on the operation of data-based markets. Accordingly, while the GDPR might indeed increase the competitiveness of European firms in privacy-enhancing technologies—the only market the Impact Assessment directly relates to with regard to comparative advantages of European firms—this is an extremely small part of all the product and service markets affected by the GDPR. Recognizing such effects might have led to a more rigorous discussion of the inherent tension between data-based innovation and data protection, and to the incorporation of some solutions. Furthermore, even within the limited framework adopted by the Commission for its analysis, the impact of the change in the legal regime on the ability of firms to realize data-based advantages should have been analyzed. This is because the GDPR, while introducing much-needed clarity and harmonization into data protection, also significantly increased the costs and difficulty of data collection and sharing due to its scope of prohibitions, sanctions, and enforcement mechanisms, thereby significantly affecting competition and innovation.

Finally, let us relate to the argument that introducing more competition implies that personal data will be held by more market players, and that realizing data synergies might increase the quality of information regarding data subjects held by such players. Yet the justifications for the GDPR do not aim at preventing wider use of data, but rather attempt to give users more control over their personal data. Moreover,

---

[166] Commission Staff Working Paper, *Impact Assessment Accompanying the GDPR*, SEC(2012)72 final, 25 January 2012, see in particular annex 10 on the Impacts of the Preferred Option on Competitiveness.
[167] *Ibid*., p. 150
[168] *Id.*
[169] *Id.*

the fact that more firms hold more or better personal information does not automatically imply harm to well-being. Rather, much depends on the use of such information. Many uses of private information may significantly increase well-being.[170] But more importantly, the question is whether we prefer that a few very large firms hold such information, or that more firms have access to such information and compete over data-based advantages. In line with economic teachings, in most cases competition can bring about benefits to users, including stricter self-imposed obligations on how the data owner will use the data. Competition will reduce the ability of large and diversified firms to use their market power and market position to obtain user consent.[171]

VI. CONCLUSION AND THE WAY FORWARD

The GDPR is the Magna Carta of data protection, the importance of which cannot be overstated. Data protection creates an inherent clash with competing values, most importantly the potential loss of benefits from better data-based knowledge.[172] To balance these tradeoffs, the GDPR does not prohibit data collection or sharing. Rather, it provides mechanisms for control, accountability and liability over data collection, processing and use, by combining individual rights with systemic governance.[173]

Yet, as this article shows, the price of data protection through the GDPR is much higher than previously recognized.[174] As elaborated, the GDPR has two main harmful effects: it limits competition in data markets, creating more concentrated market structures and entrenching the market power of those who are already strong; and it limits data synergies, thereby preventing the creation of some data-based knowledge. Such effects belie the confidence expressed by European Commissioner for Justice, Consumers and Gender Equality Věra Jourová, according to whom "the big guys increasing market share? I don't believe [the GDPR] will have such a consequence."[175]

In a world where "those that know how to use [data] have a decisive competitive advantage... through raising performance, offering more user-centric products and services, fostering innovation,"[176] and where the battlefield over data-based ad-

[170] For such a suggestion see Goldfarb and Tucker, *supra* note 148. For an example of undeniable benefits see A. R. Miller and Catherine Tucker, *Can Healthcare Information Technology Save Babies*? 119(2) JOURNAL OF POLITICAL ECONOMY 289 (2011)(privacy protections are associated with 320 annual deaths of US-born babies); Jin-Hyuk Kim & Liad Wagman, *Screening Incentives and Privacy Protection in Financial Markets: A Theoretical and Empirical Analysis*, 46 RAND J. OF ECON. 1 (2015)(loan applicants in counties in the San Francisco Bay Area that set privacy as a default paid higher interest rates, due to the reduced ability of banks to match applicants to loans).

[171] Campbell et al., *supra* note 7.

[172] Acquisiti *et al*., *supra* note 13.

[173] *Ibid*; Margot E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability* 92(6) SOUTHERN CALIFORNIA L. REV. 1529 (2019).

[174] Some commentators foresaw some of the effects of privacy laws on competition. See, e.g., Campbell et al., *supra* note 7. Yet, to our knowledge, many of the dynamics exposed in this paper have not been previously recognized.

[175] Sam Schechner and Nick Kostov, *Google and Facebook Likely to Benefit from Europe's privacy crackdown*, WSJ (April 23, 2018), https://www.wsj.com/articles/how-europes-new-privacy-rules-favor-google-and-facebook-1524536324.

[176] EU Commission, Enter the Data Economy, 21 EPSC STRATEGIC NOTES, January 11, 2017, at 1.

vantages has become global, the GDPR's effects on data markets cannot be disregarded. Two points are worth stressing. First, the GDPR has implications well beyond the geographic borders of the EU. This is because many international firms which operate in the EU, or trade with it, must comply with its rules.[177] Once such firms adopt the internal mechanisms necessary for GDPR compliance, these may be used for non-EU data as well. In addition, some jurisdictions are following in the footsteps of the EU and adopting laws which resemble the GDPR.[178] Widespread adoption of such laws could lead to even greater concentration of firms in international markets. Second, most of the effects analyzed in this article are long-term ones, which will not disappear once the market adjusts to the existence of the GDPR. It is thus worth reevaluating the overall welfare effects of the legal data regime chosen. Note that this is not to say the overall welfare effects of the GDPR are necessarily negative. The GDPR may negatively affect competition but still be welfare-enhancing. This will be the case if the harm to data subjects reduced by the GDPR is sufficient to compensate for its competitive effects, including its potential to increase users' participation in the market based on increased trust.[179] The overall balance depends on the relative magnitude of these effects.

While an in-depth exploration of tools for creating a more welfare-enhancing equilibrium is beyond the scope of this article, we offer several suggestions. Most importantly, competition law should give more weight to factors which might balance the negative effects of the GDPR on competition and innovation. For example, when evaluating the competitive effects of a merger or a joint venture, more weight should be given to considerations such as the ability of firms to engage in welfare-enhancing data sharing which may facilitate reductions in market concentration, or the potential for significant data synergies that could not be realized otherwise. This implies, for example, a more lenient policy towards mergers or joint ventures between small or medium-sized data controllers, which would enable them to reach economies of scale and scope in data analysis and compete more effectively with those that already enjoy such economies. It also implies that when at least one data controller in a proposed merger or joint venture already possesses strong comparative advantages in data analysis, a careful balance is required between the benefits of increased data synergies and the need to ensure the ability of other firms to effectively compete, in light of the increased hurdles to data collection and processing resulting from the GDPR. The conditions for applying the essential facilities doctrine and granting access to data might also need to be redefined in light of the effects of the GDPR. The interface

---

[177] ANU BRADFORD, THE BRUSSELS EFFECT (2020), chapter 5; Batikas et al., *supra* note 10.

[178] E.g., California Consumer Protection Act 1st Jan 2020, Cal AB 375 § 1798.125. The definition for the sale of data to which the restrictions of this Act apply is wider than the one adopted in the GDPR. This implies that the dynamics we identify may apply to a wider set of contexts. The Californian Act also forbids firms from refusing to deal with consumers who decline to opt in, or from charging them higher prices, thereby potentially further increasing transaction costs. Interestingly, the adoption of GDPR-like laws elsewhere around the world creates both positive and negative externalities for the EU. On the positive side, broader adoption of such laws levels the playing field with regard to foreign firms' ability to realize some data-based advantages. In addition, such a development signals a global shift in attitudes toward the protection of personal data, which in turn may strengthen acceptance of the policy among the EU's own citizens and firms. On the negative side, it might increase international concentration and limit some data-based innovations that might have also benefitted EU consumers.

[179] See also Campbell *et al.*, *supra* note 7.

between the GDPR and competition law, in cases where harm to privacy is minimal and benefits to competition and innovation are large, may also need to be reevaluated.[180]

In addition, assessments of market power and potential competition should take into account the actual competitive effects of the GDPR. No longer can it be assumed that new players seeking to accumulate large volumes of data face only low barriers, as was the case in several earlier Commission decisions,[181] especially where separate entities collect different parts of the dataset. Consider, for example, the Commission's reasoning when it approved a joint venture between Google and the global biopharmaceutical firm Sanofi aimed at using big data analysis to improve the management and treatment of diabetes.[182] The competitive analysis at the time addressed concerns that the venture would allow the parties to lock in patients by restricting their ability to direct their data towards alternative services. The Commission dismissed such claims on the grounds that data subjects had a right to data portability. Yet, as shown in this article, such a right is generally not sufficient to address competitive concerns. This is because the comparative advantages of the joint venture are partially based on existing large datasets owned by the parties. It may be difficult for competitors to overcome such comparative advantages unless they can convince a sufficiently large number of users to sign up to their services, or unless they can combine different datasets. The first option is limited by user "stickiness," by the fact that the data potentially arrives in fragmented form and at different points in time, and by the fact that economies of scale and scope and network effects in data analysis create significant first-mover advantages that are difficult for new competitors to overcome. The second option, data sharing, must overcome high hurdles, as elaborated in this article.

We offer several other suggestions that might go some way toward making the GDPR more welfare-enhancing. First, where uncertainty regarding how to meet the GDPR's legal obligations contributes to concentration, it may be useful to consider ways of limiting such uncertainty. For instance, as elaborated above, the GDPR establishes a right to data portability, but does not specify technical requirements for meeting this commitment. The development (by regulators together with industry) of technological standards for data portability and interoperability might help reduce the consequent uncertainty as to what standards might ultimately be applied.[183]Second,

---

[180] See also Nicolo Zingales, *Data Protection Considerations in EU Competition Law: Funnel or Straightjacket for Innovation*? in THE ROLE OF INNOVATION IN COMPETITION ANALYSIS (Paul Nihoul and Peter Van Cleynenbreugel eds., 2018)(Recognizing that the fundamental right to data protection cannot be ignored by competition enforcers, the author calls for a framework of cooperation between competition and data protection authorities).

[181] For example, in the case of the Facebook/WhatsApp merger, the Commission ruled that "…there will continue to be a large amount of Internet user data that are valuable for advertising purposes and that are not within Facebook's exclusive control." Case M.7217 – Facebook/ WhatsApp, Commission decision pursuant to Article 6(1)(b) of Council Regulation No 139/2004, para.189. In the Microsoft/LinkedIn case, the Commission found that "the combination of their respective datasets does not appear to result in raising the barriers to entry/expansion for other players in this space, as there will continue to be a large amount of internet user data that are valuable for advertising purposes and that are not within Microsoft's exclusive control." Case M.8124 – Microsoft / LinkedIn.

[182] COMP/M.7813 SANOFI/GOOGLE/DMI JV, rec 63 et seq. 4. See also COMP/M.7337 IMS HEALTH/CEGEDIM BUSINESS, rec. 218.

[183] Gal and Rubinfeld, *supra* note 38.

governments might support investment in the development of better and privacy protection tools which can retain more value from collected data.[184] Third, priority could be given to the development of better and faster tools for verification of GDPR compliance. And fourth, certification of data management and vetting processes could go a long way toward reducing costs. The government can either certify such tools or help facilitate such certification.[185] In addition, the use of certified tools should be taken into account when assessing liability, and presumptions based on the use of reasonable tools should be created. This, in turn, could significantly reduce the risks involved in data sharing.

A final suggestion relates to the structuring of mandatory data-sharing obligations under other laws in a way which is sensitive to the fact that it has become more difficult for small or new firms to grow and enjoy significant data synergies by obtaining data from external sources. To illustrate, under the Public Sector Information Directive,[186] some types of governmental data must be shared. The regulation does not differentiate between sharing with firms that already possess much data and with those that do not. As argued by De la Mano and Padilla in another context, this may entrench the dominance of the former to the extent that economies of scale exist in the analysis of such data.[187] Our findings expose additional grounds for this effect, especially in situations where governmental data can act as a partial substitute for personal data. It might thus be worth considering the option of asymmetric sharing of data, so that in certain circumstances the obligation to share data will relate (mainly) to sharing it with small or new entities. In line with this suggestion, it is worth exploring whether more flexible mechanisms for obtaining user consent, such as opt-out rather than opt-in, should be applied with regard to certain types of data, the benefits of which are undeniable.[188]

As this paper shows, privacy policy is interlinked with competition and the resultant data-based innovation in more ways than have yet been recognized. In particular, the GDPR raises the transaction costs of sharing data between different data controllers. We hope that recognizing such effects will lead to reevaluating the balance reached and to the adoption of tools to ensure that overall welfare is increased. While measuring privacy harms and comparing them to competition and innovation harms is an extremely challenging task, which extends beyond this paper, the solutions we suggest generally avoid the need for such careful balancing, while assuming that data protection is an essential factor in well-being.

---

[184] See also Layton, *supra* note 117, p. 11-12.
[185] ISO certification is already available for some data collection activities.
[186] Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the Re-Use Of Public Sector Information Text with EEA relevance.
[187] De la Mano and Padilla, *supra* note 19.
[188] *See, e.g.*, Goldfarb and Tucker, *supra* note 148; Miller and Tucker, *supra* note 170.

Appendix: Questions asked in interviews with firms whose business model is based on the utilization of private data:

1. Based on your experience, what effects do you think that the GDPR (and similar regulation) imposed on M&A activities? Did they strengthen, weaken, or not affect incentives to merge in order to utilize the data of the target firm?
2. Has the GDPR significantly changed the ability to make use of private data gathered by a potential target firm before the merger?
3. Are these short-term or long-term obstacles? In other words, is this just a transition period and data-motivated mergers will soon flourish again?
4. What, if any, are the tools that you use—or that you know others in the industry use—in order to overcome such legal obstacles to the use of private data collected by the target firm?
5. Has this led to changes in market interactions- for example do you see more internal growth? More joint ventures, etc.?