

# Data as a Common Good

Essays on Data Portability and B2B Industrial Data Sharing

Dissertation submitted to attain the degree of

*Doctor rerum politicarum (Dr. rer. pol.)*

at the School of Business, Economics and Information Systems  
of the University of Passau

Submitted by: Nadine Stüdlein, M.Sc.  
First Reviewer: Prof. Dr. Jan Krämer  
Second Reviewer: Prof. Dr. Thomas Widjaja

April 2022



**Accepted as dissertation**

at the School of Business, Economics and Information Systems of the University of Passau

**Date of the disputation:** July 7, 2022

**Chair of the examining committee:** Prof. Dr. Markus Diller

**First reviewer:** Prof. Dr. Jan Krämer

**Second reviewer:** Prof. Dr. Thomas Widjaja

# Abstract

Data is an important resource in our economy and society, substantially improving overall business efficiency, innovativeness and competitiveness, and shaping our everyday lives. Yet, to leverage the data's full potential, its access and availability is vital. Thus, data sharing across organizations is of particular importance. This thesis examines the role of data sharing in the digital economy and contributes to a better understanding why data sharing matters, why it is still underutilized, and how data sharing can be encouraged. Thereby, the thesis contributes to the ongoing academic debate as well as the practical and political efforts on how to promote data sharing.

The thesis is comprised of three studies. Study 1 examines personal data sharing among (competing) online services. Particularly, it investigates the consequences of Article 20 in the General Data Protection Regulation (GDPR, May 2018), ensuring the right to data portability. This relatively new right allows online service users to transfer any personal data from one service provider to another. Focusing on a) the amount of data provided by users and b) the amount of user data disclosed to third party data brokers by service providers, the study investigates the right to data portability's effect on competitiveness and consumer surplus. Study 2 and Study 3 focus on non-personal data sharing among competing firms. Study 2 examines the literature to identify and classify barriers to non-personal, machine-generated data sharing. The study explains firms' reluctance to sharing data and discusses policy and managerial implications for overcoming the data sharing barriers. Study 3 focuses on data sharing via platforms. It investigates the Business-to-Business (B2B) data sharing platform design implications for promoting industrial data sharing. In particular, Study 3 investigates the dimensions control and transparency regarding their effect in eliciting cooperation and encouraging data sharing among firms.

In summary, this thesis examines and reveals how access and availability of data can be increased through creating beneficial data sharing conditions in B2B relationships. Particularly, the thesis contributes to the understanding of a) the implications of data sharing laws, defined in the GDPR for personal data and b) the challenges and measures of the not yet successfully established, non-personal data sharing.

# Acknowledgments

First and foremost, I would like to thank Prof. Dr. Jan Krämer, who supervised me excellently throughout the whole time and supported me in any situation. I really appreciate having a supervisor who is an excellent researcher and cares about my work, but at the same time about me as a person. Furthermore, I would like to thank Prof. Dr. Thomas Widjaja for being the second examiner of my thesis, and Prof. Dr. Markus Diller who agreed to be the chairperson of the examining committee.

Many thanks also go to my colleagues and friends at the University of Passau. First, I would like to thank Dr. Daniel Schnurr for all the technical and organizational support and also for his help with any other questions and concerns. Second, I would like to thank my co-author and friend Dr. Oliver Zierke for the great collaboration in research projects and for the many joint activities. Moreover, I gratefully thank all my other team colleagues over the years, Dr. Victoria Fast, Dr. Bastian Haberer, Janina Hofmann, Alexander Witte, Nikolai Sachs and Andreas Schauer, whom I could always count on, and Katrin Gleixner for her organizational support at the chair.

Finally, I would like to thank my parents Gerhard and Ingrid for all their support throughout my life, without whom this thesis would not have been possible, and Kim for his love, support and encouragement, as well as my little son Jannik, who puts a smile on my face every day.

# Contents

<b>Introduction</b>	<b>1</b>
<b>Data portability, data disclosure and data-induced switching costs: Some unintended consequences of the General Data Protection Regulation</b>	<b>10</b>
Introduction . . . . .	11
The model . . . . .	12
Equilibrium analysis . . . . .	12
Comparison of market outcomes . . . . .	13
Conclusions and limitations . . . . .	14
Appendix . . . . .	14
References . . . . .	15
<b>Developing a Framework for Strategic Data Sharing Barriers among Competitors</b>	<b>16</b>
Introduction . . . . .	17
Defining the Concept of Data Sharing . . . . .	20
Barriers to Data Sharing . . . . .	22
Policy and Managerial Implications . . . . .	41
Conclusions and Directions for Future Research . . . . .	48
References . . . . .	49
<b>Sharing needs Caring: Experimental Insights on the Optimal Design of B2B Data Sharing Platforms</b>	<b>66</b>
Introduction . . . . .	68
Related Literature . . . . .	70
Experimental Design . . . . .	76
Procedures . . . . .	79
Data Analysis and Results . . . . .	81
Discussion . . . . .	85
Conclusions and Directions for Future Research . . . . .	87
References . . . . .	89

# Introduction

Data is essential to our economy and society and increasingly impacting our lives. Data enables personalized products and services, targeted advertising, individualized healthcare and is essential for developing new technologies, such as Internet of Things (IoT) and Artificial Intelligence (AI), allowing to improve overall business efficiency, innovativeness and competitiveness (European Commission 2020, Krämer et al. 2020, Ministry of Economic Affairs and Climate Policy 2019, Jones and Tonetti 2020). However, in order to benefit from data effectively, data access and availability is key. In essence, data needs to be shared between firms to leverage their full potential (Jarke et al. 2019, Martens et al. 2020, De Prieëlle et al. 2020, Koutroumpis et al. 2020).

Politicians already intend to encourage data sharing to increase data access and availability, aiming to create an ecosystem in Europe to boost the data economy to match its economic weight by 2030. Particularly, the European Commission aims at creating a unified data space to facilitate data access and availability for both personal and non-personal data (European Commission 2020). Despite these intentions and efforts, data sharing is still extensively underutilized and rarely established effectively, inhibiting the potential of the data. In turn, understanding and promoting data sharing is an important avenue for companies, researchers, and policymakers alike. This thesis targets this highly relevant and important research topic through the lens of Information Systems Research, by combining an economic, legal and technical perspective of data sharing.

By conceptualizing data as a common good and analyzing the implications for data sharing in a Business-to-Business (B2B) context, this thesis contributes to understanding the contemporary hindrances for data sharing in industries and across service providers. Thus, it links closely to the ongoing debate on data availability, particularly B2B data sharing, by revealing how to encourage data sharing. This research thereby contributes to the vivid discussions both in the political and academic spheres.

Figure 1 illustrates the focus of the thesis by showing the researched parties and their relationships: There are two companies, A and B. These companies are in competition with each other and data is shared by company A with company B. Study 1 focuses on sharing personal data between (competing) online services. Thereby, the data can either be transferred directly from A to B, or it is first ported to the user, to whom the data belongs to. Study 2 and Study 3 focus on sharing non-personal data among competing industries; the data can again be shared either directly (Study 2) or via a platform, which acts as an intermediary between the businesses (Study 2 and Study 3).

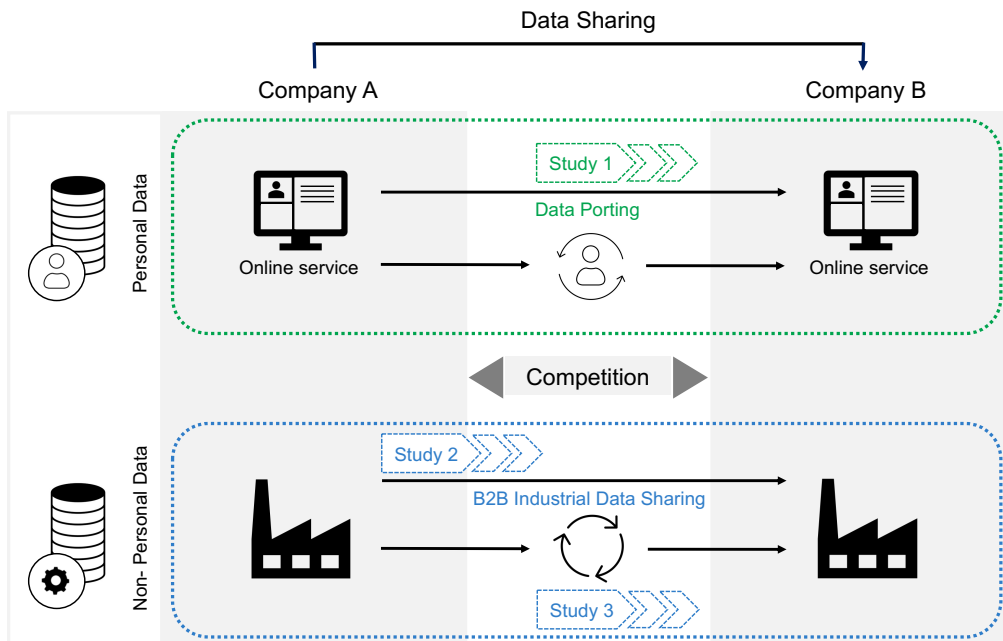


Figure 1: Illustration of the scope of this thesis and its three individual studies.

In summary, this thesis shows how data access and availability can be increased by altering and shaping prerequisites to B2B data sharing as well as understanding the consequences of data sharing. Thereby, this thesis contributes to the understanding of a) the implications of data sharing laws, defined in the General Data Protection Regulation (GDPR) for personal data and b) the problems and measures of the not yet established non-personal data sharing. The following paragraphs offer a brief summary of the thesis' central findings across the three studies.

**Study 1** (Krämer and Stüdlein 2019) examines personal data sharing among (competing) online services and thereby investigates the consequences of Article 20 in the GDPR (May 2018), ensuring the right to data portability. This relatively new right allows online service users to transfer any personal data from one service provider to another. Thereby, data is either provided by the online service to the user or transferred directly between online services. The study is motivated by the European Commission's intention to facilitate switching between content providers (CPs) on the one hand and increase market entry of new providers on the other (c.f., e.g., European Commission 2017). Yet, it is not well researched how the right to data portability affects users' privacy, service usage prices, and consumer surplus. Additionally, while the right to data portability increases competition, it is not clear how it affects content providers profitability and competitiveness. Focusing on a) the amount of data provided by users and b) the amount of user data disclosed to third party data brokers by service providers, this study investigates some of the unintended consequences of the right to data portability.

In the light of scarce empirical data, this study develops a game-theoretic model for data sharing and particularly accounts for the right to data portability. The model assumes that content providers can generate revenue from the price users pay for using their service and by selling users' data to third parties. Thus, content providers benefit from more users in two ways: More revenue from the use of the service and more data, entered by users on the platform. Users, on the other hand, have a gross utility from using the service, but also face a disutility through several costs, including privacy costs, costs for using the platform, preference costs, and data-induced transaction costs (e.g., costs for providing the data). The more data a user provides, the higher its gross utility, but also the higher the privacy costs for the share of data that a content provider discloses to third parties.

The model assumes a setting in which an incumbent content provider initially enjoys a monopoly position, but is faced by a new (competing) content provider, who enters the market in the second period. This is especially relevant because it shows how the right to data portability affects the switching decision of consumers who were active with the incumbent content provider in the previous period as well as the pricing and data disclosure decision of the providers. In each period, first the content provider/s announce/s its data disclosure level, that is the share of personal data they disclose to third parties; and second, the users decide if and which content provider they use. In the second period, the model differentiates between *old* consumers, that have used the incumbent CP before and *new* consumers who have not. The case for which data portability becomes relevant is the one in which some users who used the incumbent CP in the first period switch to the new CP in the second period. With data portability, the content provider's choice in the second period no longer depends on how much data a user shared in the first period as there are no data-induced switching costs anymore.

The findings show that without data portability, the incumbent CP strategically discloses less user data in the first period in order to make users reveal more data. This leads to increased switching costs for *old* consumers in the second period and thus to a competitive advantage for the incumbent CP. Consequently, the right to data portability leads to more data disclosure by the incumbent to third parties than without data portability, resulting in higher privacy costs for users and therefore to less user data provision. Taking into account different disclosure levels and transaction costs with and without data portability, the findings also show that with data portability, the incumbent CP sets lower prices in both periods, while the entrant CP demands higher prices. Overall, the incumbent CP can no longer take advantage of switching costs, reducing its profitability. For new CPs, on the other hand, the right to data portability is beneficial, because they can set higher prices and thus increase their profitability.



The consequences for consumer surplus (and provider's competitiveness) are different for different user groups. For old consumers who used the incumbent's service before and who decide to stay with the incumbent's service, the lower price overall leads to higher utility for consumers and thus increases the incumbent's competitiveness. Old consumers who switch to the new entrant CP also benefit from their decision, otherwise they would not have switched providers. Finally, new consumers using the new entrant CP's service in the second period reduce their utility, due to higher prices; essentially decreasing the competitiveness of the entrant CP.

**Study 2** (Stüdlein 2022) surveys the literature, identifies and classifies barriers to B2B industrial data sharing among competitors and develops a conceptual framework explaining why B2B data sharing is rarely established. Implications for strategic decision-making are analyzed to provide recommendations for policymakers and managers interested in incentivizing B2B data sharing and enhancing data access and availability. The study contributes to the ongoing debate on promoting sharing of non-personal, machine-generated data, the importance of which is repeatedly emphasized by politicians (European Commission 2020) and researchers (Koutroumpis et al. 2020, De Prieëlle et al. 2020, Martens et al. 2020). However, to effectively promote data sharing, it is important to first understand what is preventing companies from sharing their own data with other businesses. Yet, the ongoing debate lacks a systematic understanding of factors that prevent firms from sharing their data, especially among competitors. This is a major shortcoming, as it offers few suggestions on how more companies participate in data sharing and how data can be shared more extensively, ultimately failing to realize its potential for businesses and society (Jarke et al. 2019).

Summarizing and systematizing the existing literature, practitioner reports, policy proposals, and guidelines offers a comprehensive overview of the increasing number of studies and helps to create a systematic understanding of factors preventing data sharing among firms. The framework comprises legal, economic, and technical barriers and analyses the economic implications. Since corporate decision-making behavior is central in B2B data sharing, and economic challenges have not yet been studied in depth, this study mainly focuses on economic implications that directly influence the corporate decision-making calculus of firms. To understand the economic calculus, the identified barriers are linked to the strategic organizational challenges of unawareness, uncertainty, and incapability. Additionally, moderators of the economic barriers are identified and analyzed. Finally, based on this framework, policy and managerial implications are developed to assist in overcoming the barriers to effectively promote data sharing.

The findings show that economic barriers pose formidable barriers to data sharing in three ways:

First, lacking expertise in data collection, analyzing and sharing raises organizational unawareness and obscures data sharing opportunities, increasing the complexity and cost of data sharing. Second, the high complexity, the resulting uncertainty in effective monetizing and competitive performance and the subsequently arising uncertainty and costs in contracting data sharing agreements lower returns on investments and hinder organizations in enacting data sharing opportunities. Third, liability risks and potential extensive hidden liability costs in case of violations, can render organizations incapable to share data in an economically compliant manner, proving to be a show stopper to data sharing.

Overall, this paper provides a holistic overview of economic challenges companies face when sharing data. The link to the strategic problems of unawareness, uncertainty, and incapability allows a prioritization of the issues, contributing to an understanding of how to effectively promote B2B data sharing. While unawareness about data sharing is a challenge that can be largely resolved within an organization, for example, by providing additional resources in the form of staff, time, or funding to explore the topic of data sharing, uncertainty is more challenging. However, uncertainty can largely be resolved within data sharing networks and thus within interorganizational relationships such as industry, business, and trade associations. The strategic problem of incapability, on the other hand, requires, for example, additional legislation directly addressing data sharing and is thus, the most challenging one.

**Study 3** (Krämer, Stüdlein, and Zierke 2021) focuses on data sharing via platforms using an economic laboratory experiment. The experiment investigates how adapting the design of B2B data sharing platforms can foster non-personal data sharing. In particular, Study 3 analyzes the dimensions control and transparency in their effect on eliciting cooperation and encouraging data sharing among firms.

Motivated by a number of existing data sharing initiatives in practice, this study investigates the impact of the widely varying platform designs. There are initiatives where all participating firms have access to all shared data, such as the *Open Data Initiative* by Adobe, Microsoft, and SAP (Microsoft 2019), and there are other initiatives where companies can share selected data with selected partners, such as the *Data Intelligence Hub* by Deutsche Telekom (Telekom 2019). Essentially, how data is shared via platforms can be characterized along two dimensions: control and transparency. *Control* is the degree to which a firm can control other firms' data access, and *transparency* is the degree of information a firm has about other firms' sharing behavior and the resulting profits. However, all these initiatives fall short of their expected impact. This study therefore examines whether, how, and to what extent the design parameters of control and transparency can be conducive to more data sharing.

Based on a literature review, it is assumed that more control in the form of the ability to punish, leads to more cooperation and thus more data sharing (c.f., e.g. Cinyabuguma et al. 2005, Güth et al. 2007, Maier-Rigaud et al. 2010, Roux and Thöni 2015). For transparency, the literature suggests that the effect of transparency depends on environmental factors (Ostrom 2000, Møllgaard and Overgaard 2001) and thus, no clear effect can be derived for the data sharing context.

In the experiment, three levels of control were considered. At the lowest level (referred to as *Collective Sharing*), a firm can control how much data it provides on the platform, but cannot control who accesses it and to what extent. At the medium level (*Excludable Sharing*), a firm can exclude other firms from accessing its data provided on the platform, and at the third level (*Individual Sharing*), a firm can individually decide how much data it makes accessible for each sharing partner.

Regarding the dimension of transparency, two levels were considered. First, *Full Information*, under which each firm can observe every previous data sharing transaction and resulting profits by all other firms, even if the firm itself was not a transaction partner in the sharing relationship. Second, *Partial Information*, under which each firm only has knowledge on previous transactions in which it was actively involved, but with no information on transactions that involve the other firms only. Along these dimensions, the experiment investigates which data sharing environment are more conducive to elicit cooperation and data sharing by firms.

The results confirm that control is necessary to foster overall data sharing. However, a medium level of control is already sufficient to increase a firm's propensity to share data on a B2B data sharing platform and additional control does not further increase data sharing. In other words, firms need control over which other firms have access to its data to share data at all, but there is no significant increase in data sharing if firms can additionally discriminate the scope or scale of access for individual firms.

Further, the results show that transparency is an important antecedent for encouraging data sharing, even though no consistent effect can be derived from the literature. In the context of data sharing, transparency of all data transactions on a B2B sharing platform fosters confidence in the system as a whole, as well as in other sharing partners, because it allows to observe in particular a sharing partner's behavior across all sharing relationships. However, as confidence particularly evolves over time, the study also shows, that with a medium and high level of control, and for both levels of transparency, more data is shared in long-lived market relationships than in short-lived ones.

In summary, the study shows that both control and transparency independently leads to more data

sharing. Thus, it suggests that platforms can partially substitute control with transparency, or the other way around to improve overall data sharing. Nevertheless, the study concludes that access control and transparency are two important dimensions that should be considered when designing B2B data sharing platforms.

To conclude, this thesis contributes to the emerging research stream on data sharing by conceptualizing data as a common good and investigating impacts and implications for data sharing and utilization. The common good conceptualization thereby deepens the understanding of data sharing, yet it also shows that data sharing does not only provide benefits but also comes at distinct costs. This nuanced understanding provides important implications for effective data sharing, both to policymakers and managers in data-driven markets alike.

Three central findings of this thesis are particularly striking and essentially comprise the core contribution to research. First, switching between online service providers and encouraging new providers market entry is considered desirable, however, personal data sharing reduces the profitability of incumbent content providers and the utility of some consumer groups. Second, while B2B industrial data sharing becomes increasingly relevant to our economy and society, costs for firms arise but have not yet been sufficiently considered. Legal, economic and technical barriers directly affect the decision-making calculus and providing an answer to the question why B2B data sharing is still underutilized. And third, in data sharing initiatives where firms lack control over managing data access to their shared data and lack information on other firms sharing behavior, less data is shared overall. Thus, to realize the full potential of data as a common good, these costs should be considered and priced accordingly when introducing future policies and practical projects to promote data sharing in the digital economy. Overall, the thesis highlights the critical impact of balancing transparency and control effectively to enable data sharing. So far, the ongoing debate on data access and availability overemphasizes the benefits and partially neglects the costs arising in data sharing. By analyzing the decision-making calculus, current hindrances, and consequences of data sharing in industries and across service providers the thesis reveals how to encourage data sharing in general. Nevertheless, different market environments and the individual behavior of firms should be analyzed in more detail. Therefore, this thesis should be understood as one building block of the ongoing debate and calls for further research.

Three potential avenues for further research appear particularly relevant and fruitful: First, future research should analyze which sharing environments and tools are appropriate to mitigate the identified problems and costs associated with data sharing. Although this thesis offers recommendations to mitigate potential barriers to data sharing, there is little empirical research on specific solutions and

thus a gap between theoretical governance structures and empirical mechanisms remains. Krämer, Stüdle, and Zierke (2021) investigate control and transparency experimentally, yet further research analyzing additional concepts as well as analyzing the effects of mechanisms implementation outside a laboratory setting is needed. Second, because of the great economic and societal importance of data sharing, future research could take a closer look at subjects' attitudes and decisions to share data. For example, subjects' risk perceptions of privacy or trade secrets, their attitudes toward uncertainty, and their trust in various potentially competing organizations that receive data requires further studies. Third, future research should analyze data sharing using complementary methods. Particularly, with regard to a subject's decision to share or not to share data, case studies and interviews on data sharing and use might be interesting, exploring the opportunities and challenges involved. Field experiments are vital in this endeavor to strengthen external validity and prove whether the suggested solutions obtain the desired results in practice.

## Bibliography

- Cinyabuguma, Matthias, Talbot Page, Louis Putterman. 2005. Cooperation under the threat of expulsion in a public goods experiment. *Journal of Public Economics* **89**(8) 1421–1435.
- De Prieëlle, Fabian, Mark De Reuver, Jafar Rezaei. 2020. The role of ecosystem data governance in adoption of data platforms by internet-of-things data providers: Case of dutch horticulture industry. *IEEE Transactions on Engineering Management* 1–11.
- European Commission. 2017. Fact Sheet: Questions and Answers - Data protection reform package. Available at [http://europa.eu/rapid/press-release\\_MEMO-17-1441\\_en.pdf](http://europa.eu/rapid/press-release_MEMO-17-1441_en.pdf). Accessed January 2, 2022.
- European Commission. 2020. Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions "a european strategy for data". Communication COM(2020) 66 final, European Commission, Brussels.
- Güth, Werner, M Vittoria Levati, Matthias Sutter, Eline Van Der Heijden. 2007. Leading by example with and without exclusion power in voluntary contribution experiments. *Journal of Public Economics* **91**(5-6) 1023–1042.
- Jarke, Matthias, Boris Otto, Sudha Ram. 2019. Data sovereignty and data space ecosystems. *Business & Information Systems Engineering* **61**(5) 549–550.
- Jones, Charles I, Christopher Tonetti. 2020. Nonrivalry and the economics of data. *American Economic Review* **110**(9) 2819–2858.
- Koutroumpis, Pantelis, Aija Leiponen, Llewellyn DW Thomas. 2020. Markets for data. *Industrial and Corporate Change* **29**(3) 645–660.

- Krämer, Jan, Daniel Schnurr, Sally Broughton Micova. 2020. The role of data for digital markets contestability: case studies and data access remedies. [https://cerre.eu/wp-content/uploads/2020/08/cerre-the\\_role\\_of\\_data\\_for\\_digital\\_markets\\_contestability\\_case\\_studies\\_and\\_data\\_access\\_remedies-september2020.pdf](https://cerre.eu/wp-content/uploads/2020/08/cerre-the_role_of_data_for_digital_markets_contestability_case_studies_and_data_access_remedies-september2020.pdf). Centre on Regulation in Europe (CERRE) Policy Report.
- Krämer, Jan, Nadine Stüdlein. 2019. Data portability, data disclosure and data-induced switching costs: Some unintended consequences of the general data protection regulation. *Economics Letters* **181** 99–103.
- Krämer, Jan, Nadine Stüdlein, Oliver Zierke. 2021. Data as a public good: Experimental insights on the optimal design of b2b data sharing platforms. Working Paper.
- Maier-Rigaud, Frank P, Peter Martinsson, Gianandrea Staffiero. 2010. Ostracism and the provision of a public good: experimental evidence. *Journal of Economic Behavior & Organization* **73**(3) 387–395.
- Martens, Bertin, Alexandre de Streel, Inge Graef, Thomas Tombal, Néstor Duch-Brown. 2020. Business-to-business data sharing: An economic and legal analysis. <https://www.econstor.eu/bitstream/10419/227701/1/1726726282.pdf>. JRC Technical Report. Working Paper.
- Microsoft. 2019. The open data initiative. <https://dynamics.microsoft.com/en-us/open-data-initiative/>. Accessed January 2, 2022.
- Ministry of Economic Affairs and Climate Policy. 2019. Dutch digitalisation strategy: Dutch vision on data sharing between businesses. <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&ved=2ahUKEwj2kd0qmc3jAhVFD0wKHdieBb4QFjADegQIBBAC&url=https%3A%2F%2Fwww.government.nl%2Fbinaries%2Fgovernment%2Fdocuments%2Freports%2F2019%2F02%2F01%2Fdutch-vision-on-data-sharing-between-businesses%2FDutch%2Bvision%2Bon%2Bdata%2Bsharing%2Bbetween%2Bbusinesses.pdf&usg=AOvVaw10VaXErN1EgRhDZD8h19Gi>. Accessed January 2, 2022.
- Møllgaard, H Peter, Per Baltzer Overgaard. 2001. Market transparency and competition policy. <https://www.econstor.eu/bitstream/10419/208446/1/cbs-wp2001-06.pdf>. Working Paper.
- Ostrom, Elinor. 2000. Collective action and the evolution of social norms. *Journal of Economic Perspectives* **14**(3) 137–158.
- Roux, Catherine, Christian Thöni. 2015. Collusion among many firms: The disciplinary power of targeted punishment. *Journal of Economic Behavior & Organization* **116** 83–93.
- Stüdlein, Nadine. 2022. Developing a framework for strategic data sharing barriers among competitors. Working Paper.
- Telekom. 2019. Data intelligence hub. <https://dih.telekom.net/en/>. Accessed January 2, 2022.

# Data portability, data disclosure and data-induced switching costs: Some unintended consequences of the General Data Protection Regulation

**Authors** Jan Krämer, University of Passau, Germany.  
Nadine Stüdlein, University of Passau, Germany.

**Published in**<sup>1</sup> Economics Letters (VHB-JQ3 rating: *B*).

## **Abstract**

*We develop a simple game-theoretic model to demonstrate that with the new General Data Protection Regulation's (GDPR) right to port data between content providers (CPs), (i) the incumbent CP has less incentives to preserve users' privacy, (ii) a new entrant CP will charge higher prices for its service, and (iii) customers of the new CP are worse off, while customers of the incumbent CP are better off.*

---

<sup>1</sup>Reprinted as part of the formal submission of a dissertation (<https://www.elsevier.com/about/policies/sharing>) and for personal use as defined by the Elsevier journal author rights (<https://www.elsevier.com/about/policies/copyright>). This paper is available on ScienceDirect: <https://doi.org/10.1016/j.econlet.2019.05.015>.



Contents lists available at ScienceDirect

Economics Letters

journal homepage: [www.elsevier.com/locate/econlet](http://www.elsevier.com/locate/econlet)



# Data portability, data disclosure and data-induced switching costs: Some unintended consequences of the General Data Protection Regulation



Jan Krämer\*, Nadine Stüdlein

Chair of Internet and Telecommunications Business, University of Passau, Dr.-Hans-Kapfinger-Str. 12, 94032 Passau, Germany

## HIGHLIGHTS

- We develop a two-period model to model consumer switching between content providers.
- Switching costs are constituted by the effort to re-enter data at the new provider.
- The right to data portability under the GDPR drives switching costs to zero.
- This would increase the incumbent's incentives to disclose user data in the first period.
- While customers of the incumbent are better off, customers of the entrant are worse off.

## ARTICLE INFO

### Article history:

Received 20 March 2019  
Received in revised form 6 May 2019  
Accepted 8 May 2019  
Available online 11 May 2019

### JEL classification:

L13  
L51  
L86

### Keywords:

General Data Protection Regulation  
Data portability  
Switching costs  
Privacy

## ABSTRACT

We develop a simple game-theoretic model to demonstrate that with the new General Data Protection Regulation's (GDPR) right to port data between content providers (CPs), (i) the incumbent CP has less incentives to preserve users' privacy, (ii) a new entrant CP will charge higher prices for its service, and (iii) customers of the new CP are worse off, while customers of the incumbent CP are better off.

© 2019 Elsevier B.V. All rights reserved.

## 1. Introduction

In the EU the right to port data between content providers (CPs) has recently been introduced in Article 20 of the General Data Protection Regulation (GDPR) in an effort to facilitate users' switching of CPs and to increase the competitiveness of online markets. Building upon the literature on switching costs (Klemperer, 1987, 1988), as well as on competition in privacy (Casadesus-Masanell and Hervas-Drane, 2015; Lefouilli and Toh, 2018), we develop a simple two-period model, where an incumbent CP is a monopolist in the first period and competes against an entrant CP in the second period. We highlight that *without* the possibility to port data, the incumbent CP will disclose *less* user data in the first period, because in this vein the

CP can induce consumers to provide more data. In turn, this leads to higher data-induced switching costs in the second period. However, with users' right to port data costlessly, this incentive of the incumbent vanishes. As a result, the incumbent CP discloses more user data in the first period, and the competitive position of the entrant in the second period is strengthened. Therefore, the incumbent's price is lower and its consumers are indeed better off under the new regulation. However, the entrant's price is higher and its consumers are strictly worse off.

In related work Wohlfarth (2019) also considers a two-period set-up, but does not consider that consumers' data revelation depends on the CP's service quality as well as on privacy costs; nor can CPs set a disclosure level strategically. Instead, CPs only have one strategic variable, the price charged to consumers, which is interpreted as the CP's amount of collected user data. Ultimately, this leads to the conclusion that the right to data portability will *reduce* the amount of data that is collected by the incumbent

\* Corresponding author.

E-mail addresses: [jan.kraemer@uni-passau.de](mailto:jan.kraemer@uni-passau.de) (J. Krämer), [nadine.stuedlein@uni-passau.de](mailto:nadine.stuedlein@uni-passau.de) (N. Stüdlein).



CP in the first period. In contrast, in our set-up CPs have two strategic variables, price and disclosure level, and our results show that more user information is revealed. However, in line with Wohlfarth (2019), we find that under the right to data portability the incumbent's first period price is lower.

Similarly, in the context of data portability Lam and Liu (2018) also consider a two-period model with entry in the second period. However, the authors neither consider data disclosure and privacy costs, nor the strategic role of prices. Rather they focus on competition through quality improvements by data analytics. More specifically, they assume that data analytics gives rise to learning and network effects. In this vein data portability can hinder switching to new entrants, because it would diminish the learning and network effects and thereby reinforces the lock-in effect. Consequently, in contrast to our set-up, in which entrants are better off with data portability, the additional consideration of data analytics leads to the result that data portability makes it more difficult for the entrant to acquire users. However, in line with Lam and Liu (2018), we find that data portability increases consumers' data provision in the first period.

## 2. The model

The incumbent CP *A* enjoys a monopoly position in period  $t = 1$ , but faces competition by CP *B* in period  $t = 2$  (see Klemperer (1988) for a similar set-up). In each period there exists a unit mass of the same consumers with preference  $y \sim U[0,1]$ . The CPs are horizontally differentiated, with CP *A* being located at  $y = 0$  (in periods  $t = 1, 2$ ) and CP *B* being located at  $y = 1$  (only in  $t = 2$ ).

CPs have two streams of revenue. First, CP  $i$ ,  $i \in \{A, B\}$  can demand a price  $p_{i,t}$  in period  $t$  from users for access to its service. Second, by using the service of  $i$ , consumers reveal an amount of personal data in each period  $t$ , denoted by  $x_{i,t}$ . We assume that CPs can sell this data in the data market for a marginal revenue of  $r$ . However, in order to protect users' privacy, firms can choose to sell only a fraction  $d_i$  of the entire data available to them. In summary, CP  $i$ 's profit in period  $t$  is given by

$$\Pi_{i,t} = D_{i,t} (r d_{i,t} x_{i,t} + p_{i,t}), \quad (1)$$

where  $D_{i,t}$  denotes the mass of consumers using CP  $i$ 's service in period  $t$ . Consequently, CP *A* makes a total profit of  $\Pi_A = \Pi_{A,1} + \Pi_{A,2}$  and CP *B* a total profit of  $\Pi_B = \Pi_{B,2}$ .

A consumer  $y$ 's utility when choosing CP  $i$  in period  $t$  is given by

$$U_{i,t}(y, x_{i,t}) = V(x_{i,t}) - d_{i,t} \theta x_{i,t} - p_{i,t} - \tau_i(y) - C_{i,t}(x_{i,t}), \quad (2)$$

where  $V(x_{i,t})$  is the gross utility from using the service;  $\theta > 0$  denotes a consumer's marginal privacy costs, for which we assume that  $\theta < r$  in order to focus on cases where selling consumer data is a viable business model for the CP;  $\tau_i(y)$  are the preference costs, given by  $\tau_A(y) = \tau y$  and  $\tau_B(y) = \tau(1 - y)$  for CP *A* and *B*, respectively; and  $C(x_{i,t})$  are data-induced transaction costs. These elements of the utility functions are described in more detail below.

**Consumers' information revelation and privacy costs.** We assume that consumers determine endogenously how much data,  $x_{i,t}$ , they reveal to CP  $i$  in period  $t$  (see Casadesus-Masanell and Hervas-Drane (2015) and Lefouili and Toh (2018) for a similar approach). Thereby, consumers consider two effects. First, providing more data renders the service more useful for a consumer. Specifically, we assume that the gross utility of CP  $i$ 's service is given by  $V(x)$ , with  $V(0) = 0$ ,  $\frac{\partial V}{\partial x} > 0$ ,  $\frac{\partial^2 V}{\partial x^2} < 0$ . For expositional clarity, we adopt  $V(x_{i,t}) = v x_{i,t} - x_{i,t}^2$  in the following. Second, revealing more data leads to higher privacy costs, depending on the

CP's disclosure level  $d_i$ . For example, a music streaming provider may ask a consumer to indicate which songs she likes or dislikes. The more information is provided, the more accurate the music recommendations for this consumer. However, this information may also be sold to third parties, giving rise to privacy costs.

**Transaction costs and switching costs for providing data.** In addition consumers are assumed to face transaction costs,  $C_{i,t}$  when providing data to a CP. For example, consumers may find it cumbersome to recall which songs they liked and disliked when asked by a new music streaming service, even if privacy issues are not of concern. Consumers have to bear such transaction costs for every new piece of information they provide. Thus, in  $t = 1$ , irrespective of the regulatory regime,  $C_{A,1} = c x_{A,1}$  for all consumers of CP *A*. In the second period, one must differentiate between "old" customers, denoted by superscript  $o$ , that have used CP *A* in  $t = 1$ , and "new" customers, denoted by superscript  $n$ , who have not. Without data portability, data-induced transaction costs are given by  $C_{A,2}^o = c \max\{x_{A,2} - x_{A,1}, 0\}$  for old customers staying with CP *A* and in all other cases  $C_{i,2} = c x_{i,2}$ . With data portability, the only difference is that  $C_{i,2}^o = c \max\{(x_{i,2} - x_{A,1}), 0\}$ , i.e., old customers face the same transaction cost in the second period, irrespective of whether they stay with CP *A* or switch to CP *B*. As in Klemperer (1987), we assume that consumers are myopic and do not anticipate the entry of CP *B* in  $t = 2$ .

**Timing.** In period  $t = 1$ , CP *A* first announces its disclosure level,  $d_{A,1}$ , and price,  $p_{A,1}$ ; subsequently consumers decide whether they use CP *A*'s service and how much information  $x_{A,1}$  they provide. Thereafter, in period  $t = 2$ , CP *A* and CP *B* first simultaneously announce their disclosure levels,  $d_{A,2}$  and  $d_{B,2}$ , and prices,  $p_{A,2}$  and  $p_{B,2}$ ; subsequently consumers decide which CP they use and how much information  $x_{A,2}$  or  $x_{B,2}$  they provide.

We solve for the subgame perfect Nash-equilibrium with and without data portability and differentiate equilibrium outcomes, where necessary, by superscript  $P$  and  $NP$ , respectively.

## 3. Equilibrium analysis

### 3.1. No data portability

**Users' decision in  $t = 2$ .** A consumer of CP  $i$  in  $t = 2$  would optimally reveal  $x_{i,2}^* = \arg \max_x U_{i,2} = 1/2(v - c - d_{i,2} \theta)$  information. Note that the optimal amount of information revelation depends only on CP  $i$ 's disclosure level in that period, and is independent of prices or whether the consumer has previously revealed information to CP *A* in  $t = 1$ . However, as is shown in the following, a user's CP choice in  $t = 2$  does depend on her previous choice of CP, individual preference and prices. In order to focus on the most interesting case we assume that not all consumers have chosen CP *A* in  $t = 1$ . Assuming market coverage in  $t = 2$ , the location of the indifferent old and new consumers, for which respectively  $U_{A,2} = U_{B,2}$  holds, is given by<sup>1</sup>

$$\tilde{y}_{n,2} = \frac{(p_{B,2} - p_{A,2}) + (x_{A,2}^{*2} - x_{B,2}^{*2}) + \tau}{2\tau} \quad (3)$$

$$\tilde{y}_{o,2} = \tilde{y}_{n,2} + \frac{x_{A,1}^* c}{2\tau}. \quad (4)$$

Clearly,  $\tilde{y}_{o,2} > \tilde{y}_{n,2}$  for  $c > 0$ , because old customers do not have to re-enter their data when staying with CP *A*. Let  $\tilde{y}_1 = D_{A,1}$  denote CP *A*'s market share in  $t = 1$ , then three different cases have to be considered (see Fig. 1).

<sup>1</sup> The assumption that the market is not covered in the first period, but covered in the second period, models a situation in which entry leads to a market expansion. For this to occur, the exogenous model parameters, specifically,  $\tau$ ,  $v$ ,  $\theta$  and  $c$  must be at intermediate values.

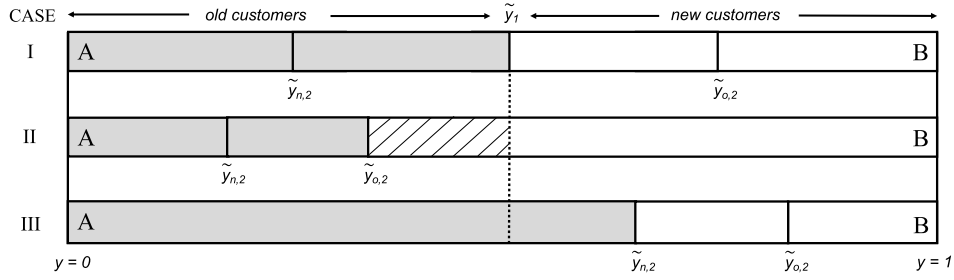


Fig. 1. Distribution of indifferent old and new consumers in  $t = 2$  without data portability.

In case I, i.e.,  $\tilde{y}_{n,2} < \tilde{y}_1 < \tilde{y}_{o,2}$ , all old customers would stay with CP A and all new customers would go to CP B. Thus, there is effectively no competition for users, as marginal changes in a CP's price or disclosure level have no impact on its second period demand. Because no old user of CP A switches to the new CP B, data portability is irrelevant in this case. In case II, i.e.,  $\tilde{y}_{n,2} < \tilde{y}_{o,2} < \tilde{y}_1$ , some of the first period users of CP A switch to CP B in the second period. Thus, CPs do compete for users in prices and disclosure levels and data portability matters. In case III, i.e.,  $\tilde{y}_1 < \tilde{y}_{n,2} < \tilde{y}_{o,2}$ , no first period user of CP A switches to CP B in the second period. Instead, in the second period CP A extends its user base to some of the new users. Thus, data portability is irrelevant again in this case.

Clearly, case II is the most interesting with respect to the strategic role of data portability, and hence we focus on this case in the following. In case II, the total demand of old and new customers for CP A and B is given by  $D_{A,2}^* = \tilde{y}_{o,2}$  and  $D_{B,2}^* = 1 - \tilde{y}_{o,2}$ .

Note that without data portability, the indifferent old customer in  $t = 2$ , and thus a CP's market share, depends on  $x_{A,1}^*$ , i.e., the amount of information that the old customer has provided to CP A in  $t = 1$  (which again depends on A's strategic choice of the disclosure level in  $t = 1$ ). This is what creates *data-induced switching costs* for old customers in period  $t = 2$ , which can then be exploited by CP A in period  $t = 1$ .

**CPs' decision in  $t = 2$ .** Substituting  $x_{i,2}^*$ ,  $D_{A,2}^*$  and  $D_{B,2}^*$  into CP  $i$ 's second period profit and simultaneously solving the first order conditions  $\frac{\partial \Pi^{A,2}}{\partial p_{A,2}} \stackrel{!}{=} 0$ ,  $\frac{\partial \Pi^{A,2}}{\partial d_{A,2}} \stackrel{!}{=} 0$ ,  $\frac{\partial \Pi^{B,2}}{\partial p_{B,2}} \stackrel{!}{=} 0$ ,  $\frac{\partial \Pi^{B,2}}{\partial d_{B,2}} \stackrel{!}{=} 0$  yields:

$$d_{A,2}^* = d_{B,2}^* = \frac{(v-c)(r-\theta)}{(2r-\theta)\theta} \quad (5)$$

$$p_{A,2}^* = \frac{-3r^3(v-c)^2 - 8r\theta^2(x_{A,1}^*c + 3\tau) + 2\theta^3(x_{A,1}^*c + 3\tau)}{6\theta(-2r+\theta)^2} + \frac{r^2\theta(3v^2 - 6vc + 8x_{A,1}^*c + 3c^2 + 24\tau)}{6\theta(-2r+\theta)^2} \quad (6)$$

$$p_{B,2}^* = \frac{-3r^3(v-c)^2 + 8r\theta^2(x_{A,1}^*c - 3\tau) - 2\theta^3(x_{A,1}^*c - 3\tau)}{6\theta(-2r+\theta)^2} + \frac{r^2\theta(3v^2 - 6vc - 8x_{A,1}^*c + 3c^2 + 24\tau)}{6\theta(-2r+\theta)^2} \quad (7)$$

**Users' decision in  $t = 1$ .** In the first period, customers reveal  $x_{A,1}^* = \arg \max_x U_{A,1} = 1/2(v-c-d_{A,1}\theta)$  information to CP A. Consequently, CP A can influence  $x_{A,1}^*$  only via its choice of the disclosure level  $d_{A,1}$ . The indifferent consumer  $\tilde{y}_1$  ( $U_{A,1}(\tilde{y}_1) = 0$ ), and thus, CP A's demand in  $t = 1$  is given by  $D_{A,1} = \tilde{y}_1 = \frac{x_{A,1}^{*2} - p_{A,1}}{\tau}$ .

**CP A's decision in  $t = 1$ .** CP A maximizes its total profit  $\Pi_A = \Pi_{A,1}(p_{A,1}, d_{A,1}) + \Pi_{A,2}(d_{A,1})$  over both periods by setting  $p_{A,1}^*$  and  $d_{A,1}^*$ .<sup>2</sup>

### 3.2. With data portability

**Users' decision in  $t = 2$ .** Again, we focus on the interesting case II, where some of the old CP A customers switch to CP B in  $t = 2$ . With data portability, an old customer's CP choice does not depend on how much information she has revealed in  $t = 1$  anymore, i.e.,  $\tilde{y}_{o,2} = \tilde{y}_{n,2}$  as in (3). It follows that  $D_{A,2}^* = \tilde{y}_{n,2}$  and  $D_{B,2}^* = 1 - \tilde{y}_{n,2}$ .

**CPs' decision in  $t = 2$ .** Due to the absence of data-induced switching costs, both CPs are now fully symmetric in  $t = 2$  and face the same strategic trade-offs. Thus, they will set the same prices and disclosure levels of

$$p_{A,2}^* = p_{B,2}^* = \frac{r^2(v-c)^2(-r+\theta)}{2\theta(-2r+\theta)^2} + \tau \quad (8)$$

$$d_{A,2}^* = d_{B,2}^* = \frac{(v-c)(r-\theta)}{(2r-\theta)\theta} \quad (9)$$

**Users' decision in  $t = 1$ .** Since users are myopic, they face the same strategic trade-off in  $t = 1$  as in the case without data portability.

**CP A's decision in  $t = 1$ .** Considering only profits from  $t = 1$ , CP A's optimal price and disclosure level are

$$p_{A,1}^* = -\frac{r^2(v-c)^2(2r-3\theta)}{8\theta(-2r+\theta)^2} \quad (10)$$

$$d_{A,1}^* = \frac{(v-c)(-r+\theta)}{\theta(-2r+\theta)} \quad (11)$$

## 4. Comparison of market outcomes

### 4.1. Disclosure levels and amount of user-revealed data

Intuitively, in the second period no CP can use the disclosure level strategically to create data-induced switching costs in subsequent periods. Hence, irrespective of whether data portability is possible or not, both CPs face the same strategic trade-offs with respect to the disclosure level in  $t = 2$ , and thus second period disclosure levels are always identical for both CPs in both regimes (cf. (5) and (9)), i.e.,  $d_{i,2}^{NP} = d_{i,2}^P$  for  $i \in \{A, B\}$ .

Moreover, recall that in case I and III, data portability is irrelevant for the strategic interaction. Hence, CP A's disclosure level in the first period must also be unaffected by data portability and is readily given by (11). However, in case II it can be shown that the disclosure level of CP A in  $t = 1$  must in fact be higher with data portability than without data portability (see Appendix). The intuition for this result is as follows. Without data portability CP A strategically lowers the disclosure level in the first period in order to induce users to reveal more data. This increases the data-induced switching costs of the old customers in the second period, which can then be used to mitigate competition between CP A and B.

<sup>2</sup> The closed form solutions are lengthy and omitted here.

**Proposition 1.** *Under the right to data portability, the incumbent content provider is (weakly) willing to disclose more user data to third party data brokers than without data portability, and, therefore, users provide (weakly) less data.*

#### 4.2. Prices

First, recall that with data portability the CPs compete on equal footing and will set the same disclosure level and price (cf. (8)) in the second period. Second, notice that without transaction costs ( $c = 0$ ), the second period prices with and without data portability coincide. Third, from Eqs. (6) and (7) it can be seen that for positive transaction costs ( $c > 0$ ) and under our assumption that  $r > \theta$ , it follows that  $p_{A,2}^{NP} > p_{i,2}^P > p_{B,2}^{NP}$ . Indeed, it follows from  $\frac{\partial p_{A,2}}{\partial d_{A,1}} = -\frac{c\theta}{6} < 0$  that setting a lower disclosure level in the first period, induces CP A to demand a higher price in the second period. Finally, in the Appendix it is shown that absent data portability, the incumbent CP A will also set a higher price in the first period, i.e.,  $p_{A,1}^{NP} > p_{A,1}^P$ .

**Proposition 2.** *Under the right to data portability, the incumbent content provider demands lower prices in both periods than without data portability. However, the price of the entrant content provider is higher with data portability.*

#### 4.3. Profits and consumer surplus

It is evident that the incumbent CP makes less profits under the right to data portability, because it cannot exploit the presence of switching costs in the second period. Moreover, it is also evident that the right to data portability increases the profits of the entrant CP, because it can charge higher prices. With respect to consumer surplus, it is worthwhile to consider the effects separately for the different groups of customers. First, old consumers using CP A in both periods are better off with the right to data portability: In the second period, they experience a lower price, everything else being equal; and in the first period, they also experience a lower price, but a higher disclosure level. It can be shown that overall utility in the first period is higher with data portability when  $v$  is sufficiently large (see Appendix). Second, old customers switching from CP A to CP B must therefore also be better off, because otherwise they would not have switched providers. Third, new customers of CP B in the second period must be worse off. This is because the price of CP B is higher with the right to data portability, while the disclosure level remains constant.

**Proposition 3.** *Under the right to data portability, the incumbent CP makes lower profits, while the entrant CP makes higher profits. Furthermore, with data portability old consumers of the incumbent CP are strictly better off iff  $v > c + d_{A,1}\theta(2r - \theta)$ . However, new consumers of the entrant CP are strictly worse off.*

### 5. Conclusions and limitations

We have offered a simple stylized model that points at some of the unintended consequences of the new right to data portability introduced as part of the General Data Protection Regulation in the EU. Specifically, we have shown that an incumbent CP may disclose more user data under this new regulation. Although the regulation increases the competitiveness of the incumbent, leaving old customers better off, it decreases the competitiveness of the entrant, rendering potential new customers worse off.

Evidently, our simple model is subject to many limitations that may be addressed by future research, such as issues of endogenous entry (allowing for entry deterrence), longer time

horizons (allowing for richer strategies), as well as a fraction of consumers being foresighted and anticipating the possibility of entry. While this will produce more nuanced results, the basic economic mechanism presented here would still be at work in all of these extensions.

### Appendix

#### Proof of Proposition 1

It remains to be shown that  $d_{A,1}^{P^*} > d_{A,1}^{NP^*}$  in equilibrium in case II. In all other cases, and in  $t = 2$ , disclosure levels will be the same with and without data portability. See that with data portability, CP A sets its disclosure level for the first period to maximize its profit in  $t = 1$ . Thus, in equilibrium we have  $\frac{\partial \Pi_{A,1}(d_{A,1}^{P^*})}{\partial d_{A,1}} = 0$ . As the profit function has a maximum at this point, it follows immediately that  $\frac{\partial^2 \Pi_{A,1}(d_{A,1}^{P^*})}{\partial d_{A,1}^2} < 0$ . Without data portability, CP A chooses its disclosure level while taking into account not only the profitability of the first period, but also that the information revelation in the first period affects its second period market share through data-induced switching costs. Therefore, CP A sets  $d_{A,1}^{NP^*}$  so that it maximizes the total profits, i.e., in equilibrium it must hold that  $\frac{\partial \Pi_{A,1}(d_{A,1}^{NP^*})}{\partial d_{A,1}} + \frac{\partial \Pi_{A,2}(d_{A,1}^{NP^*})}{\partial d_{A,1}} = 0$ . It can be shown that  $\frac{\partial \Pi_{A,2}}{\partial d_{A,1}} < 0$ , if  $\tau \geq \frac{1}{6}c\theta$ , i.e., if  $\tau$  is sufficiently large. Consequently, in order to satisfy the first-order condition (5),  $\frac{\partial \Pi_{A,1}}{\partial d_{A,1}} > 0$  must hold in equilibrium. Since the marginal profit in  $t = 1$  falls in  $d_{A,1}$  locally around the equilibrium value (see (5)), it follows that  $d_{A,1}^{P^*} > d_{A,1}^{NP^*}$ . Finally, a higher disclosure level of the service increases the privacy costs and therefore the disutility a user derives from being active at the considered CP. Consequently, it reduces the amount of data users are willing to reveal. In other words,  $d_{A,1}^{P^*} > d_{A,1}^{NP^*}$  implies that users are willing to provide less data under the right to data portability than without data portability, i.e.,  $x_{A,1}^{P^*} < x_{A,1}^{NP^*}$ .  $\square$

#### Proof of Proposition 2

It remains to be shown that  $p_{A,1}^{NP^*} > p_{A,1}^{P^*}$  in equilibrium in case II. See that with data portability, CP A sets its price for the first period to maximize its profit in  $t = 1$ . Thus, in equilibrium we have  $\frac{\partial \Pi_{A,1}(p_{A,1}^{P^*})}{\partial p_{A,1}} = 0$ . Without data portability, CP A chooses its prices while taking into account also the profitability of the second period. Therefore, CP A sets  $p_{A,1}^{NP^*}$  so that it maximizes the total profits, i.e., in equilibrium it must hold that  $\frac{\partial \Pi_{A,1}(p_{A,1}^{NP^*})}{\partial p_{A,1}} + \frac{\partial \Pi_{A,2}}{\partial p_{A,1}} = 0$ . Since the profit from the second period is independent of  $p_{A,1}$ , we have  $\frac{\partial \Pi_{A,2}}{\partial p_{A,1}} = 0$ . Thus, with the same disclosure level in the first period, the profit-maximizing price  $p_{A,1}^{P^*}$  would be the same with and without data portability. However, since the disclosure level of CP A in  $t = 1$  is lower without data portability and  $\frac{\partial p_{A,1}}{\partial d_{A,1}} = \frac{1}{4}(\theta(-v + c + d_{A,1}\theta) + r(-v + c + 2d_{A,1}\theta)) < 0$ , it follows that  $p_{A,1}^{P^*} < p_{A,1}^{NP^*}$ .  $\square$

#### Proof of Proposition 3

It remains to be shown that with data portability, old consumers of the incumbent CP are strictly better off with data portability. See that for any given disclosure level and ensuing equilibrium prices, consumers' utility in the first period is identical with and without data portability, i.e.,  $U_{A,1}^P(d_{A,1}) = U_{A,1}^{NP}(d_{A,1})$ .

Furthermore, it holds that  $\frac{\partial U_{A,1}(p_{A,1}^{p*})}{\partial d_{A,1}} > 0$ , if  $v > c + d_{A,1}\theta(2r - \theta)$ .

Now the Proposition follows immediately from  $d_{A,1}^p > d_{A,1}^{NP*}$  from Proposition 1.  $\square$

## References

- Casadesus-Masanell, R., Hervas-Drane, A., 2015. Competing with privacy. *Manage. Sci.* 61 (1), 229–246.
- Klemperer, P., 1987. The competitiveness of markets with switching costs. *Rand J. Econ.* 18, 138–150.
- Klemperer, P., 1988. Welfare effects of entry into markets with switching costs. *J. Ind. Econ.* 37, 159–165.
- Lam, W.M.W., Liu, X., 2018. Does data portability facilitate entry? Norwich Business School and Centre for Competition Policy Working Paper. Available at [https://editorialexpress.com/cgi-bin/conference/download.cgi?db\\_name=EARIE45&paper\\_id=253](https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=EARIE45&paper_id=253).
- Lefouili, Y., Toh, Y.L., 2018. Privacy regulation and quality investment. Toulouse School of Economics Working Paper 17 (795).
- Wohlfarth, M., 2019. Data portability on the internet: An economic analysis. *Bus. Inf. Syst. Engrg.* 1–24. <http://dx.doi.org/10.1007/s12599-019-00580-9>, articles in advance.

# Developing a Framework for Strategic Data Sharing Barriers among Competitors

**Author** Nadine Stüdlein, University of Passau, Germany.

**Submitted to** Information Systems Frontiers (VHB-JQ3 rating: B).

## **Abstract**

*Sharing of non-personal, machine-generated data among businesses holds great potential for our economy and society, because Business-to-Business (B2B) data sharing promotes business innovation and competitiveness. However, B2B data sharing is rarely established. While firms consider it beneficial to use other firms' data, they are reluctant to share their own. While extant literature emphasizes the value of data sharing among companies, there are strikingly few studies that examine the barriers to data sharing and especially barriers among competitors. To address this gap, the paper identifies and classifies data sharing barriers among competing firms and develops a framework explaining the reluctance to sharing data. The framework reveals legal and technical barriers, and it particularly emphasizes the economic barriers, linking them to strategic problems; namely unawareness, uncertainty, and incapability. Based on this framework, guidelines for policy-makers and managers are developed for identifying, addressing, and overcoming the barriers and to effectively promote data sharing.*

# Developing a Framework for Strategic Data Sharing Barriers among Competitors

Nadine Stüdlein\*

## **Abstract**

Sharing of non-personal, machine-generated data among businesses holds great potential for our economy and society, because Business-to-Business (B2B) data sharing promotes business innovation and competitiveness. However, B2B data sharing is rarely established. While firms consider it beneficial to use other firms' data, they are reluctant to share their own. While extant literature emphasizes the value of data sharing among companies, there are strikingly few studies that examine the barriers to data sharing and especially barriers among competitors. To address this gap, the paper identifies and classifies data sharing barriers among competing firms and develops a framework explaining the reluctance to sharing data. The framework reveals legal and technical barriers, and it particularly emphasizes the economic barriers, linking them to strategic problems; namely unawareness, uncertainty, and incapability. Based on this framework, guidelines for policymakers and managers are developed for identifying, addressing, and overcoming the barriers and to effectively promote data sharing.

*Keywords:* B2B data sharing, barriers, strategic challenges, corporate decision-making

## **1 Introduction**

Data, regularly labeled the “gold of the 21st century”, is a key resource to business performance and an essential antecedent to and driver for innovation (OECD, 2015; European Commission, 2011; Kerber, 2016). Data allows companies to create personalized products and services, enables targeted advertising, and is vital to using and advancing new technologies, such as Internet of Things (IoT) and Artificial Intelligence (AI). Thereby, data can substantially improve overall business efficiency, innovativeness and competitiveness (European Commission, 2020a; Krämer et al., 2020; Chen et al., 2014; Hall, 2020). Yet, leveraging

---

\*Chair of Internet and Telecommunications Business, University of Passau, Dr.-Hans-Kapfinger-Str. 12, 94032 Passau, Germany. Email: nadine.stuedlein@uni-passau.de. Declarations of interest: none

data effectively is challenging, as it requires extensive and diverse data sets, for which internal company data rarely suffices; making data sharing between businesses vital to fully exploit the potentials promised (Jarke et al., 2019; Wixom et al., 2020; De Prieëlle et al., 2020; Ministry of Economic Affairs and Climate Policy, 2019; Jernigan et al., 2016; Koutroumpis et al., 2020).

However, Business-to-Business (B2B) data sharing is rarely established, hindering extensive data exploitation, despite its economic potential (European Commission, 2020a; Otto, 2019; Richter and Slowinski, 2019). Over 75 percent of companies solely analyze in-house data and do not trade with other companies (Barbero et al., 2018).<sup>1</sup> And among those companies sharing data, data is shared very selectively and often only to an incremental extent (Arnaut et al., 2018b). Even in the selective cases where data is shared between firms, 80% of it relates to customer data (Elsaify and Hasan, 2021). This leads to an overall under-utilization, especially of non-personal, machine-generated (industrial) data, resulting in neglected business opportunities and missed chances for successful innovation (Martens et al., 2020). The under-utilization of data becomes even more striking when considering that efforts for collecting already existing data represent sunk costs to businesses, and thus even extensive data sharing only bears little additional costs (Martens et al., 2020). Ultimately, welfare gains cannot be leveraged, leaving enormous potential benefits to businesses, consumers, and society untapped.

Prominent examples illustrating the economic and societal benefits of data sharing can be found among car manufacturers and traffic providers, allowing companies to develop intelligent mapping services. The mapping service HERE, supported by the competing car manufacturers Daimler, BMW and Audi, does not only create additional company profits, but also enhances road safety and transportation sustainability (HERE, 2021a). Through sharing vehicle sensor data from cars on a neutral platform, HERE provides real-time traffic information, to effectively route drivers or alert them regarding dangerous situations (HERE,

---

<sup>1</sup>Based on the actual behaviour of 100 European companies implementing big data solutions as sampled by Deloitte in a recent study. Sub-contractors, e.g., for analyses are excluded.

2021b; McLoughlin, 2021; Arnaut et al., 2018a). At the same time, car manufacturers can monetize their vehicle data via the HERE marketplace and generate additional profits when the data is used in developing additional services (HERE, 2021c). Another prominent example can be found in health care, where clinics share their data to develop machine learning algorithms that enhance cancer diagnoses and treatment (Cammarota et al., 2020; Ngiam and Khor, 2019). This benefits society through better medical services (Rutella et al., 2020) and at the same time benefits the clinics economically through quicker, more accurate and more cost effective diagnoses (Kerr et al., 2020).

Understanding and overcoming barriers to data sharing therefore is a key challenge for companies, researchers, and policymakers alike (European Commission, 2020a; Pauer et al., 2018; EBA Open Banking Working Group, 2018). However, extant literature on barriers to data sharing is surprisingly scarce, providing little insights for researchers and decision-makers and inhibiting data sharing. This is a major shortcoming, as it offers few suggestions on promoting companies data sharing and data sharing extent, ultimately failing to realize the data's potential for businesses and society (Jarke et al., 2019). To fill this void and to improve data sharing the existing barriers and inhibitors to data sharing are identified and analyzed.

This paper therefore examines the obstacles companies face in data sharing and analyzes the underlying barriers. Thereby it particularly focuses on B2B data sharing with competitors, as this provides a complex and especially challenging setting for data sharing and thus provides a comprehensive context for identifying data sharing barriers. Building upon the identification and illustration of the interrelated barriers, this paper develops a framework to structure the barriers in data sharing and to guide policymakers, managers, and future research in improving data sharing.

The remainder of this paper is organized as follows: In Section 2 the relevant terminology and a conceptual categorization for classifying B2B data sharing between competitors are introduced. Based on this classification, in Section 3, the extant economic literature on



barriers to B2B data sharing is reviewed and linked to three empirically identified strategic problems in data sharing. Next, Section 4 discusses the contribution to research and derives policy and managerial implications. Finally, Section 5 concludes by summarizing the main results and identifying possible limitations of this work and extensions to research.

## **2 Defining the Concept of Data Sharing**

In light of the rapidly advancing technology and in order to provide accurate definition of data sharing, it is necessary to explicate the concepts of data and data sharing and to briefly introduce the underlying literature relevant to the article.

Strikingly, the extant literature offers no homogeneous definition of data. Building upon the definition by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), *data* can be understood as “a reinterpretable representation of information in a formalized manner, suitable for communication, interpretation or processing” (European Commission, 2014, p.4).<sup>2</sup> As such, data can be both human-generated or machine-generated in a sense that the data that is “created without the direct intervention of a human by computer processes, applications or services, or by sensors processing information received from equipment, software or machinery, whether virtual or real” (European Commission, 2017b, p. 9). Either type of data, human- or machine-generated, can be created intentionally or arise as a “by-product” of actions and interactions (European Commission, 2014). Furthermore, data can distinguished into personal or non-personal data, depending on whether it contributes to the identification of a natural person or not (European Commission, 2017b; GDPR.EU, 2021a). This aspect is particularly relevant, as personal data is regulated more extensively, requiring an empirical distinction based on the data content.

*Data sharing* in turn emphasizes making data available to others for use and re-use (Arnaut et al., 2018b; Michener, 2015; Tenopir et al., 2011), used “to describe all possible

---

<sup>2</sup>See also ISO/IEC 2382-1, revised by ISO/IEC 2382:2015.

forms and models underpinning B2B data access or transfer” (European Commission, 2018a, p.5).<sup>3</sup> Important characteristics of data sharing are that data has low distribution and reproduction costs and is non-rivalrous in its use.<sup>4</sup> Therefore and in contrast to physical goods, data is an ideal resource for sharing (Ministry of Economic Affairs and Climate Policy, 2019; Koutroumpis et al., 2020; Martens et al., 2020).

In the literature, B2B data sharing is further distinguished by the degree of data accessibility. Data sharing can occur privately, that is sharing exclusively among distinct partners, or publicly, in the sense that data is made openly available (European Commission, 2020a; New York Law School, 2012).

While public data sharing is determined mainly by organizational, technical and political barriers as well as legal and regulatory constraints (Dawes, 1996; Ramon Gil-Garcia et al., 2007; Gil-García and Pardo, 2005; Janssen et al., 2012), private data sharing is additionally impacted by the mutual exchange relationship and is also shaped by economic and behavioral concerns (Martens and Duch-Brown, 2020; Perkmann and Schildt, 2015; Praditya et al., 2017).

Furthermore, in private-private data sharing relations, economic aspects such as rivalry and competition become additionally relevant (European Commission, 2020a) and therefore, the literature distinguishes between *vertical data sharing*, that is data sharing along the supply chain and *horizontal data sharing*, referring to data sharing among companies operating in the same market segment (Crujssen et al., 2007).

Extant literature on private data sharing, particular in supply chain management, predominantly emphasizes vertical data sharing (Kembro et al., 2014; Kumar and Pugazhendhi, 2012; Li and Zhang, 2008). Thereby identifying especially trust and uncertainty as relevant additional characteristics determining the data sharing extent (Özer et al., 2011; Klein and Rai, 2009; Li and Lin, 2006). In contrast to vertical private data sharing, horizontal private data sharing is scarcely covered in the extant literature. An exception is sharing customer’s

---

<sup>3</sup>For distinct B2B data sharing approaches, see, e.g., Arnaut et al. (2018b).

<sup>4</sup>Non-rivalry of data: Data can be used by several firms simultaneously (Jones and Tonetti, 2020).

personal data (Zhao and Xue, 2012; Jentzsch et al., 2013; Liu and Serfes, 2006), which, however, is a very sensitive issue and often restricted by laws and regulations, such as the General Data Protection Regulation (GDPR).<sup>5</sup> Barriers in this area are therefore predominantly legal in nature.

While the extant literature is undoubtedly relevant to consumers, companies, and policymakers alike, it nonetheless covers data sharing selectively. Particularly, the literature emphasizes vertical private data sharing and the issues in sharing personal data; but omits the increasingly applicable area of private, yet non-personal, horizontal data sharing. Especially with recent technological advancements in AI and IoT, machine-generated (industrial) data sharing among competitors becomes increasingly relevant. It can drive innovation, promote product standardization and offers benchmarks, improving economic efficiency and contributing to social welfare (Caricom Competition Commission, 2021; De Prieëlle et al., 2020; Botelho, 2018; European Commission, 2020a). While industrial data is considered less critical and less strictly regulated compared to personal data, companies nonetheless remain hesitant to share their data.

### **3 Barriers to Data Sharing**

In order to explore the barriers that prevent companies from sharing data among competitors, extant literature was reviewed and specifically a wide range of practitioner reports, policies, and guidelines incorporated. This approach yielded a comprehensive range of hindrances to data sharing, originating from *legal*, *economic* and *technical problems*. Legal barriers are a relatively complex topic in its own right. From a legal stance, much has already been discussed academically and politically, and the European Commission has already been active with its *European Strategy for Data* (European Commission, 2020a), the *Data Governance*

---

<sup>5</sup>Personal data are per definition “any information relating to an identified or identifiable natural person” (GDPR.EU, 2021a, p.1).

*Act* (European Commission, 2020b), the *Data Act* (European Commission, 2022) and several Directives to facilitate B2B data sharing.

Similarly, from a technical perspective there is a vivid discussion on technical infrastructure, data sharing standards, and interoperability. For example, the European Commission is attempting to support the use of data by establishing a uniform technical infrastructure (European Commission, 2020a), and business organizations are promoting light house projects, such as the *International Data Space* (International Data Spaces Association, 2022) or *Gaia-X* (BMW, 2020). However, while this is an intriguing technical debate, to a wide range of cooperate organizations the economic implications are particularly relevant. Therefore, this paper analyzes corporate decision-making behavior and the economic implications to identify economic barriers that directly influence corporate decision-making calculus. These barriers are linked to three strategic problems in firms: *unawareness*, *uncertainty*, and *incapability*. Additionally, moderators of the economic barriers are identified and analyzed.

**Economic Unawareness** Economic barriers create the strategic challenge of unawareness, as organizations are unaware how to benefit economically from sharing the data. This illustrates a lack of expertise among companies both in sharing and in using data.

**Economic Uncertainty** Second, the strategic challenge of uncertainty in companies poses an economic barrier to data sharing. Uncertainty thereby revolves around data use by sharing partners and its potentially severe competitive and financial impacts to companies, as well as the issue of incomplete contracts.

**Economic Incapability** Finally, economic barriers to data sharing create the strategic challenge of incapability in organizations, which is triggered by liability risks and the potential extensive hidden costs in case of violations.

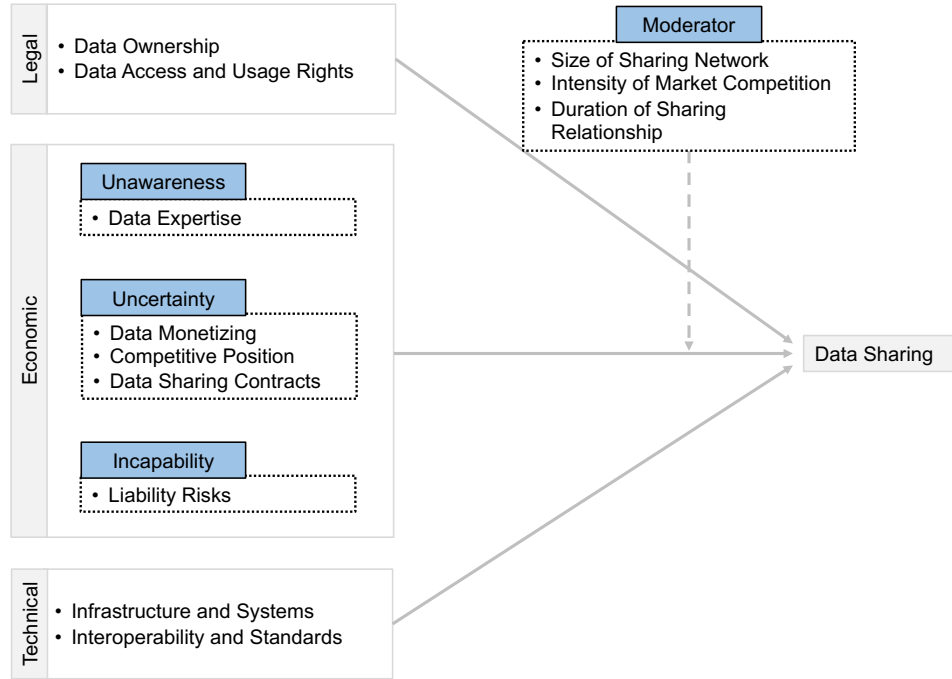


Figure 1: Barriers to B2B data sharing. Research Model.

The barriers analyzed in this paper are illustrated in Figure 1. The section is structured into the three areas of legal, economic and technical barriers. The focus of this paper is clearly on the economic ones. However, for completeness the main discussion points with respect to the legal and technical barriers are summarized in Sections 3.1 and 3.4. For the economic barriers, the respective drivers of the aforementioned strategic problems to data sharing are examined.

### 3.1 Legal Barriers

Central to triggering uncertainty in organizations from a legal stance, is the lacking consistent and comprehensive legal framework for non-personal, machine-generated data. In particular, the lacking framework poses two major hurdles to data sharing: undefined data ownership and undefined data access and usage rights. This section briefly outlines how legal barriers impact data sharing decision-making in organizations.

### **3.1.1 Data Ownership for Non-Personal Data**

There is no law specifically defining and addressing ownership rights of non-personal, machine-generated data (European Commission, 2016; Schweitzer and Peitz, 2017).<sup>6</sup> As data is an intangible and non-rivalrous good, it is also not covered by the current European legal framework for property rights (European Commission, 2017a; Asbroeck et al., 2017; Kerber, 2016). For rival goods, there is a well-developed legal framework regarding property rights (Schweitzer and Peitz, 2017) and for intangible and non-rival ideas, intellectual property rights are established (Duch-Brown et al., 2017). For the characteristics of data, intellectual property rights, such as copyright and database rights exist (Asbroeck et al., 2017); however, these do not apply to machine-generated data, as they do not arise from an intellectual effort (European Commission, 2017a). Accordingly, no such property right exists for machine-generated, non-personal data.

For companies, this leads to a strategic problem as without distinct regulation, firms are unaware or uncertain of legally compliant possibilities for data sharing. Thus, companies often do not know who owns the data, who can use the data and who holds decision rights over the data (Ministry of Economic Affairs and Climate Policy, 2019; European Commission, 2020a). Empirical studies on B2B data sharing also show that for the majority of firms, uncertainty regarding ownership rights and legally permitted application of the data is crucial (Arnaut et al., 2018b).

A right to data ownership for non-personal, machine-generated data has been extensively debated politically (European Commission, 2017a,b) and academically (Wiebe, 2017; Kerber, 2016). However, preliminary conclusions indicate that a data ownership right is neither necessary nor useful. It is argued that the introduction of a right to data ownership would

---

<sup>6</sup>For non-personal data, many legislative principles are subject to more general provisions. For example, the regulation on the free flow of non-personal data, dealing with the free movement of non-personal data across borders, the availability of data for regulatory control and data porting (European Union, 2018), or The Open Data Directive, focusing on public sector bodies and dealing with the re-usability of public sector information (European Union, 2019). Further Directives are specialized on different kinds of data, such as PSD2 Directive in case of the financial transaction data (European Union, 2015) or the Clinical Trial Regulation in case of clinical trial data (European Union, 2014).

create even larger legal uncertainties and thus inhibit exchange of data, competition and innovation (Kerber, 2016; BDVA, 2017a). For a detailed discussion about the debate of a data ownership right for non-personal data, the reader is referred to further literature, see, e.g., Kerber (2016), Schweitzer and Peitz (2017) or Martens et al. (2020).

### **3.1.2 Data Access and Usage Rights**

The debate of data ownership is essentially shifting towards data access and usage rights, which is an increasing issue in the ongoing policy debate, as companies do not necessarily have access to and usage rights for their machine-generated data (European Commission, 2019, 2018b; Barbero et al., 2018; Drexl, 2018; Martens et al., 2020). With machine-generated, non-personal data<sup>7</sup>, there is no single unique legal entity that can solely claim data access rights (Martens et al., 2020). Instead, multiple parties, such as machine manufacturers, the machine owners or the machine users are essential in data generation and possibly have interest in data access (Martens et al., 2020). Yet, manufacturers of IoT devices implement exclusive control (e.g., by encoding collected information) over data access and use (European Commission, 2018b; Drexl, 2018; Martens et al., 2020), often inhibiting small and medium sized companies (SMEs) with low bargaining power from accessing data generated in the value chain (Barbero et al., 2018). Without legal options for enforcing access to their machine-generated data organizations are effectively prevented from data sharing, despite 90% of the data generated by IoT devices remains unused (Green, 2016).

Furthermore, with a lack of defined and legally binding data usage rights, data shared by firms is hard to protect from misappropriation (Barbero et al., 2018). Access authorized firms can only use data depending on contractual agreements, which in the context of incomplete contracts exacerbates corresponding liabilities (see Section 3.2.3) and effective data monetizing (see Section 3.2.2).

---

<sup>7</sup>For personal data, the GDPR links the data to the corresponding person, who has access rights to their data regardless of the data collection.

The issues of data access and usage rights are already being discussed in detail by researchers (Schweitzer and Peitz, 2017; Martens et al., 2020; Drexl, 2018) and addressed by policymakers to enable data sharing. Contemporary attempts by the European Commission focus on strengthening access rights for machine-generated data (European Commission, 2017b), which is an important step in improving data sharing among competitors. Further, the European Commission addresses the issue of usage rights for co-generated industrial data in its recently published Data Act (European Commission, 2022).

In its detail, this is a very complex topic for lawyers and policymakers and this paper does not intend to provide a comprehensive legal framework. However, the legal debate and arising uncertainty has strong implications for organizational decision-making, which is the main emphasis of this research.

### **3.2 Economic Barriers**

Economic barriers reinforces the strategic problems of unawareness, uncertainty and incapability. Here, unawareness refers to a lack of knowledge about data and data sharing in firms, triggered by a scarcity of skilled workers present in organizations. Economic uncertainty arises from the extent in which other companies will have access to shared data and how they will use it. Organizations sharing data particularly fear jeopardizing their competitive position, and potential financial losses related to data sharing. Additionally, incomplete contracts lead to great uncertainties and high transaction costs for organizations. Finally, economic incapability is triggered by liability risks and the potential extensive hidden costs in case of violations, rendering companies economically incapable of sharing data. Economic barriers are further amplified by moderators, such as the size of the sharing network, the intensity of competition in the market and the duration of the sharing relationship, discussed in this section.



### **3.2.1 Unawareness through a Lack of Data Expertise**

An economic barrier to data sharing, related to the strategic challenge of unawareness, is the lack of data expertise. Companies, especially in the IoT domain, frequently lack the human capital to manage data projects effectively (Jernigan et al., 2016). Without distinct expertise, data collection, protection and commercialization is inhibited (Pauer et al., 2018). However, skilled workers are key in generating economic value from data (Krämer et al., 2020) and thus an important resource in data sharing. Currently, data sharing occurs primarily between companies with data science and machine-learning capabilities (Elsaify and Hasan, 2021).

In particular, many companies lack the necessary abilities and expertise to 1) evaluate the value of data (European Commission, 2017b; Barbero et al., 2018), 2) utilize their own data to generate corporate and societal value (BDVA, 2019; Arnaut et al., 2018b), and to 3) connect with data from other companies (European Commission, 2019). Additionally, maintaining abilities and expertise updated is particularly challenging, considering the rapid pace of technological advancement. Unsurprisingly, companies often fail to keep up with their organizational and individual capabilities for planning and organizing data (BDVA, 2019). To acquire and sustain the necessary human capital, additional investments in recruiting and in regular training would be essential, but many companies are reluctant to provide the substantial resources required (Barbero et al., 2018; BDVA, 2019).

This challenge is also reflected in empirical analyses. According to a study by Arnaut et al. (2018b), carried out for the European Commission, 38% of the respondent companies state that the “lack of skilled data workers inside the company” is a major obstacle to data sharing (Arnaud et al., 2018b). Thus, the lack of data expertise proves to be an economic barrier to data sharing.

### **3.2.2 Uncertainty in Effective Monetizing, Competitive Performance and Data Sharing Contracts**

Economic uncertainty inhibits data sharing, as necessary decisions and investments are not adequately supported in organizations due to unclear profit prospects and fear of losing relative competitive position. Further, costly and inefficient contracting prevents companies from sharing data.

**Data Monetization** Monetizing shared data effectively is a challenge for organizations, particularly because assessing the value of data is non-trivial (Koutroumpis et al., 2020; Arnaut et al., 2018b). Because data is an experience good, its value cannot be determined for a buyer prior to its consumption. However, after receiving the information a potential buyer is no longer willing to pay a price, resulting in a situation which Arrow labeled the Information Paradox (Arrow, 1962). As both the data holder (seller) and the data re-user (buyer) assess the value of the data differently - due to its unknown value to the re-user - prior to the transaction, data sharing is unlikely (Cazier et al., 2007).

In the case of exclusive data, for example, organizations are particularly cautious. Herein, data holders (over-)emphasize the potential benefits generated from exclusive data access in the future (Ministry of Economic Affairs and Climate Policy, 2019). Even if not able to leverage the profits (see Section 3.2.1), sharing exclusive data inhibits selling this data effectively in the future (Martens et al., 2020). However, as potential data re-users are not aware of the data, they value this data less than the data holder and are unwilling to pay an appropriate exclusivity premium (Koutroumpis et al., 2020; Arnaut et al., 2018b). Consequently, companies cannot monetize their data according to their own valuation and are therefore unwilling to hand over their data for re-use.

Moreover, the information paradox hinders data sharing partners in agreeing on models for sharing profits arising from shared data (Martens et al., 2020), especially when there are “imbalances in negotiating power” (European Commission, 2020a, p.7). Organizations cau-

tiously assess sharing partners' relative power and subsequently adapt their sharing behavior, which decisively shapes inter-organizational collaboration (Madlberger, 2008; Mäkipää, 2006). Dominated parties are afraid of disproportional profit distribution. They particularly fear that the dominant party will underestimate the value of its data and thus exploit the benefits of data sharing solely, rather than distributing benefits fairly (Childerhouse et al., 2003; Lee and Whang, 2000). Despite often not being able to exploit the value of the data on their own, organizations nonetheless fear that their sharing partners may perform decisively better and solely monetize the resulting profits. This makes organizations reluctant to cooperate and inhibits them from providing others with access to their data.

In summary, effectively monetizing data faces extensive uncertainty in organizations. As organizations themselves may not be able to exploit their data, they are exceptionally cautious in sharing exclusive data, as it is believed to be of higher value. Essentially, the different perceptions of value among data holders and re-users lead to disagreements about how to share profits arising from shared data.

**Relative Competitive Position** The relative competitive position forms a second central economic concern to organizations and a barrier to data sharing among competitors. Firms fear that data sharing erodes their competitive advantage (Jones and Tonetti, 2020; European Commission, 2020a, 2017b; Trkman and Desouza, 2012; Natwichai et al., 2005), in particular if the data is of intellectual value (Pala and Zhuang, 2019; Wixom et al., 2020). This concern arises primarily from two subordinate factors: the loss of access control and the loss of usage control through data sharing.

Access control, understood as the ability to determine who can access the data once it is shared (Ruj et al., 2011; Samarati and de Vimercati, 2000; Sandhu and Samarati, 1994), is a particularly sensitive issue in data sharing, especially when data is stored and shared via collective data sharing platforms or pools with multiple participants. In large collective data sharing initiatives, companies are insufficiently informed on other participants, who access

their shared data (Pauer et al., 2018). This reinforces companies' fear of disclosing trade secrets and fear of the impact of sharing sensitive information (such as detailed information of the production process or data preparation methods) on competition and innovation (Arnaut et al., 2018b; Barbero et al., 2018; Pauer et al., 2018; Norman, 2001; Mattioli, 2014).

The problem becomes even more striking when considering that data disclosure can destroy entire business models when it is leaked to the public, e.g., if sensitive information of the organization, such as security breaches become public (Pala and Zhuang, 2019). This would have a negative impact on customer satisfaction and thus on market value, reputation and on the firm's business volume (Pala and Zhuang, 2019; Lippert and Govindarajulu, 2006). Consequently, it is little surprising that empirical insights indicate, that being unable to assess who actually accesses data once it has been shared on a platform, is a major obstacle to sharing data at all (Pauer et al., 2018).

In contrast to access control, additionally usage control - the ability to control how companies that have been granted access, use the shared data - is a concern to organizations (Cao et al., 2020; Munoz-Arcentales et al., 2019; Pretschner et al., 2011, 2008). The data sharing organizations fear that the data will be misappropriated and used to their disadvantage (Otto, 2019; Arnaut et al., 2018b; Clemons and Hitt, 2004). These reservations are further strengthened by the non-rivalry of data and the fact that data can be easily duplicated across multiple parties, increasing behavioral uncertainties.

Companies are concerned that their data will be used to offer substitutes to their products and services rather than complementary products and services (Martens et al., 2020). Firms are exceptionally critical and reluctant to sharing data that enables substitute products and services, while they are more likely agreeing to complementary or neutral activities based on data sharing and re-use. However, data sharing provides additional information on an organization's products, services, and processes, making the organization overall more transparent to competitors. As the sharing organization is not able to control how competitors will utilize this information, increasing transparency increases also the likelihood substitutes

will emerge in their market (Martens et al., 2020) and harm the organization's profits (Trkman and Desouza, 2012). Accordingly, data sharing is inhibited by the inability to control data usage, which raises fears that sharing firms will lose their competitive advantage, its profitability, and reputation.

These concerns are further reinforced as the lack of usage control increases the vulnerability to opportunistic behavior by other companies. Opportunistic behavior is characterized by features such as cheating, distorting information, misleading partners, or offering inferior products and services (Das and Teng, 1998), and is usually triggered by distrust (Butler Jr, 1999). The more parties with individual goals use the data, the greater the potential for opportunistic behavior (Samaddar et al., 2006). This is particularly problematic for data sharing platforms with open user groups instead of closed groups where data sharing conditions are negotiated among each other (Martens et al., 2020).<sup>8</sup>

Given the lack of clear rules for responsible data use, such as legal liability (see section 3.2.3), the great uncertainty about how and for what other firms use the shared data creates competitive fears and is thus an existing hurdle hindering B2B data sharing (European Commission, 2020a; Arnaut et al., 2018b). Companies face increasing behavioral uncertainties and fear that data sharing will essentially undermine their economic activities and endanger their profits as well as their market position.

**Data Sharing Contracts** The economic barriers to data sharing are raised further by inherent problems of creating effective data sharing contracts. First, the high degree of incompleteness in such contracts is a major concern by firms. When contracting on data sharing between companies, it is usually unfeasible to list all possible rights as well as agree upon any potential scenarios in advance (Martens et al., 2020). Effective contracting is furthermore obscured as rapidly progressing technology enables different use cases as well as analyses frequently. For example, unforeseen costs that are traditionally charged to the owner of the service or good may not be attributable, as there is no legally binding definition

---

<sup>8</sup>See, e.g., Trkman and Desouza (2012) for a more detailed analysis of the potential risks.

of ownership rights in non-personal, industrial data (Martens et al., 2020). Consequently, important residual property rights, may be contracted upon incompletely, fostering immanent residual uncertainties (Asbroeck et al., 2017; Grossman and Hart, 1986). Even if comprehensive contracts, covering all relevant scenarios would be theoretically feasible, this would require extensive transaction costs for contract agreement and frequent adaptation (Duch-Brown et al., 2017; Martens et al., 2020; Schweitzer and Peitz, 2017; Asbroeck et al., 2017; Truong et al., 2012). Thus, the extensive incompleteness of contracts creates vast uncertainty and high transaction costs for organizations and constitutes a barrier to B2B data sharing.

Second, in addition to the incompleteness of contracts, bilateral agreements cannot prevent privacy and security risks sufficiently, with data being portable and easily copied (Martens et al., 2020). Since any contracts are only valid between the contracting parties, they are not enforceable against third parties, without the protection by an intellectual property right (Asbroeck et al., 2017; Duch-Brown et al., 2017; Martens et al., 2020). Unwanted and unintended data disclosure, for example through data leaks and hacking attacks, can result in data being spread publicly at an instant and cause great damage, yet contracts provide insufficient means of protection in this regard (Martens et al., 2020; Schweitzer and Peitz, 2017). The issue is even more relevant if parties to the contract cannot track whether and how data is protected, used and redistributed by the other contracting parties (Ministry of Economic Affairs and Climate Policy, 2019). The transaction risk that the receiving party will deliberately use the shared data for purposes other than those contractually agreed, to its own economic advantage and to the sharing parties disadvantage (poaching), is becoming increasingly important as a component of transaction costs (Clemons and Hitt, 2004). Ultimately, the extensive uncertainty together with the high potential risks to organizations prevent companies from making their data available, even in bilateral relationships and when contracts exist.

In summary, contracting data sharing effectively poses a major challenge to organizations. In doing so organizations face substantial (pre-contractual) transaction costs and contract

negotiations (Martens et al., 2020). However, even extensive transaction investments can nonetheless ensure sufficiently complete, secure and enforceable contracts, due to the rapid technological advancement and lacking intellectual property rights protection for industrial data, which lowers the expected value of returns. Perceiving data sharing contracting both as expensive and at the same time ineffective, inhibits and prevents data sharing.

### **3.2.3 Incapability through Liability Risks**

While for personal data, liability rules are regulated by Article 82-84 of the GDPR (GDPR.EU, 2021b,c,d), non-personal, machine-generated data lack clear and distinct definitions. This hinders a compliant distinction from comparatively extensively regulated personal data and thereby creates extensive fears of substantial liability costs in case of violations (Arnaut et al., 2018b; Ministry of Economic Affairs and Climate Policy, 2019). Depending on the type of information disclosed, data sharing can cause serious liability costs for companies and even destroy entire business models, e.g., if sensitive information, such as security breaches become public (Pala and Zhuang, 2019). This would have a negative impact on customer satisfaction, market value, reputation and business volume of the company, making companies economically incapable of sharing their data (Pala and Zhuang, 2019; Lippert and Govindarajulu, 2006).<sup>9</sup>

Another strategic problem is the unintentional disclosure of data protected under prohibitive laws, such as complementary antitrust and competition law. Such laws partially prohibiting data sharing between competing companies to prevent distortion of competition or cartelization (Ministry of Economic Affairs and Climate Policy, 2019). Thereby, companies fear that data sharing with competitors will easily lack distinction from, e.g., competition law violations. Such violations would have tremendous economic consequences for the sharing company, making organization incapable to sharing data.

---

<sup>9</sup>As of now, no regulatory intervention or contractual agreement sufficiently addresses this void (BDVA, 2017a). Civil law basically distinguishes between contractual and extra-contractual legal liability, depending on whether the liabilities are set within or outside of a contract (European Commission, 2017b).

Empirical studies have also shown that liability risks are a major barrier to data sharing (Pauer et al., 2018; Arnaut et al., 2018b). Thus, the lacking definition of liabilities and the potential extensive hidden costs make data sharing a complex challenge for organizations and leaving them even incapable to share their data.

Economic issues pose formidable barriers to data sharing in three ways: First, lacking expertise in data collection, analyzing and sharing raises organizational unawareness and obscures data sharing opportunities, increasing the complexity and cost of data sharing. Second, the high complexity, the resulting uncertainty in effective monetizing and competitive performance and the subsequently arising uncertainty and costs in contracting data sharing agreements lower returns on investments and hinder organizations in enacting data sharing opportunities. Third, liability risks and potential extensive hidden liability costs in case of violations can render organizations incapable to share data in an economically compliant manner, proving to be a show stopper to data sharing.

### **3.3 Moderators to the Economic Data Sharing Barriers**

Previously, the inherent economic barriers to data sharing were delineated. As these barriers arise in relation to the competitors' behavior, also amplifying factors were identified, which indirectly affect the extent and impact the barriers. Particular factors are 1) the size of the sharing network, 2) the intensity of competition in the market and 3) the duration of the sharing relationship. These moderators impact the amplitude of the economic barriers rather than creating additional barriers.

#### **3.3.1 Size of Sharing Networks**

An economic moderator in data sharing is the number of partners in a sharing network. The more firms are active in a network, the more data is available to any participating firms. This leads to lower participation costs (Jin et al., 2021), lower transaction costs for data access (EBA Open Banking Working Group, 2018) and increasing the effectiveness of data



sharing (Pala and Zhuang, 2019). Additionally, extensive and more diverse data provision can help foster innovation (European Commission, 2020a; Wixom et al., 2020; Jarke et al., 2019).

On the other hand, firms in larger networks and multilateral relationships experience less control over data access and data usage as well as less trust in the sharing partners (Pala and Zhuang, 2019). With more companies accessing, using and possible redistributing the provided data, the risk of single “black sheep” companies behaving opportunistically increases. Therefore, an increasing number of sharing partners increases uncertainty about the investment and commitment as it becomes less predictable what will eventually happen with the data once it is shared collectively (Ministry of Economic Affairs and Climate Policy, 2019). This fear of misappropriation and opportunistic exploitation is one of the main reasons why firms do not participate in B2B data sharing, according to the European Commission (European Commission, 2020a; Arnaut et al., 2018b).

Thus, the size of a data sharing network amplifies the economic barriers in data sharing. At the same time, a larger number of partners benefits data sharing through more detailed and thus more valuable data, while harming data sharing through increasing potential gain from opportunistic behavior.

### **3.3.2 Intensity of Competition in the Market**

Intense competition increases the incentive for sharing partners to behave opportunistically in order to strengthen their own competitive position, reducing the level of trust and cooperation in data sharing (Kollock, 1994; Rindfleisch, 2000). Therefore the economic barriers are amplified by the strategic environment of the data sharing companies (Isaac et al., 1984; Snijders and Keren, 1999).

For the sharing company, an increasing intensity of competition increases the fear of opportunistic behavior by partners, reducing incentives to data sharing. At the same time however, more competition imply more detailed data and the creation of benchmarks (De Prieëlle

et al., 2020). Because of the similarity of activities, data sharing occurs mainly between firms in the same industry (Elsaify and Hasan, 2021). The higher the (industry's) competitive level, the more similar companies are and thus the more valuable competitors' data for a company, increasing benefits and incentives for sharing (Gal-Or and Ghose, 2005).

Consequently, the intensity of competition likely impacts the economic barriers to data sharing among competitors as it increases both the value of data to a company when sharing data but also the distrust in competitors when sharing data.

### **3.3.3 Duration of Sharing Relationships**

A third moderator to the economic barriers in data sharing is the duration of data sharing relationships. Particularly **trust**, defined as: “a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another” (Rousseau et al., 1998, p.395), impacts data sharing. While longer relationships are potentially more prone to establishing trust among data sharing partners, providing deeper insights into products, services and processes over time increases a data sharing company's vulnerability to opportunism.

Short relationships with less trust in the sharing partners inhibit data sharing. Trust therein reflects expectations about partners positive motives (Das and Teng, 1998) and can be seen as a mechanism that reduces the sharing parties uncertainty (Hsu and Chang, 2014) and thus, creates and sustains cooperation between organizations (Lado et al., 1997; Özer et al., 2011; Kimmel et al., 1980). Thus, trust is a core prerequisite in data sharing (European Commission, 2018a), which can be created and strengthened through long-term interactions with repeated positive experiences (Das and Teng, 1998; Mellewigt et al., 2007). The shorter the relationship, the more difficult it is for companies to assess the partner's behavior and build positive behavioral expectations. Companies fear that partners will misappropriate the shared data beyond initial intentions and contractual agreements, such as sharing data with unauthorized third parties (European Commission, 2020a; Lee and

Whang, 2000). Consequently, data sharing is expanded with trusted partners (with long-term interactions), particularly when companies perceive the data to be sensitive (Kollock, 1994).

Furthermore, the duration of the relationship affects the expectations of **data quality**. Data quality can be assessed in terms of accuracy, timeliness and granularity (Krämer et al., 2020). Yet, these parameters are difficult to assess in advance as companies do not have information about the quality of the available data prior to a transaction (see Section 3.2.2, information paradox). However, high quality data is the basis for accurate data analysis, reliable decision-making, and the development of new technologies, such as AI (Federation of German Industries (BDI), 2020). For example, if companies provide incomplete data and subsequent procurement of the missing data is not possible or extremely expensive, the shared data can become useless (Gal and Rubinfeld, 2019). Additionally, issues around data quality, such as the validity of the shared data or punishment for bad data, are often not addressed; nor is it clear whether competing firms have an incentive to provide poor-quality or false data (SINE, 2020). Even if a company itself discloses high-quality data, it cannot be sure that it will receive high-quality data in return (reciprocity concerns), which increases uncertainty in sharing (Hsu and Chang, 2014). Empirical studies reveal that the fear of poor or insufficient data quality is a barrier to data sharing (see e.g., Arnaut et al. (2018b)). Prior experience, however, can reduce the risk and uncertainty in data quality. Thus, the relationship duration can increase relationship quality among transaction partners, increasing access to more and especially higher quality data.

Contrary to the trust building benefits of longer relationships in data sharing, deeper insights into companies increases the risk of substitutes and opportunism; potentially endangering profits and competitive advantage. Additionally, data sharing among long-term partners can inhibit new sharing relationships by imposing an intra-group bias on data sharing engagement. Nonetheless, new partners are important to create a more diverse database and thus drive innovation; particularly in dynamic markets.

In summary, the duration of data sharing relationships benefits data sharing through increasing trust and subsequently enabling sharing more detailed and higher quality data. At the same time, this detailed data can be increasingly sensitive to the organizations operations, increasing the threat through opportunistic behavior. Additionally, potential lock-in-effects in existing co-operations can hinder new partnerships and thereby under-emphasize and inhibit value generated through the additional cooperation. Thus, the relationship duration poses a third moderator in data sharing that is particularly relevant to understanding the economic barriers to data sharing.

### **3.4 Technical Barriers**

Data sharing is immanently connected to technology for accessing, analyzing and using the data and thus reporting technical barriers - in addition to legal and economic ones - may seem little surprising but important to consider. Technical challenges are indeed the most prominent obstacle for companies, reported as empirical hindrances by 73% of the companies (Arnaut et al., 2018b). However, this paper intends to emphasize the economic implications of data sharing barriers as relevant to organizational decision-makers and therefore a sophisticated IT-technical analysis is beyond its scope. Nonetheless, technical barriers are also relevant to organizations and will be analyzed in the light of their impact on decision-making. This is particularly relevant regarding a lack of technical infrastructure and advanced IT systems as well as a lack of technical specifications and standards, which limit the utility of shared data and inhibit the overall willingness to share. In particular, a lack of data interoperability and data standardization can be a show stopper for corporate data sharing initiatives.

#### **3.4.1 Lack of Technical Infrastructure and Outdated Systems**

A frequent technical barrier to data sharing originates from the technical infrastructure. With the underlying technology rapidly advancing and the demands for technological in-

infrastructure increasing with business growth and the number of products, locations and exchange relationships, technological infrastructure needs to be constantly adapted to appropriately fulfill its functions and to remain secure in doing so (EBA Open Banking Working Group, 2018; Ministry of Economic Affairs and Climate Policy, 2019). When it comes to data sharing, often an emergent concern to an organizations technological infrastructure are insufficiently available or outdated systems, inhibiting effective data sharing (EBA Open Banking Working Group, 2018). This barrier is further reinforced, as agreeing on and implementing a reliable and secure infrastructure for cross-organizational data sharing, including hardware, software, and trained staff, is highly expensive and increasingly complex for companies (Fawcett et al., 2007; Bouchbout and Alimazighi, 2008; Wixom et al., 2020; Arnaut et al., 2018b).

However, a compatible technical infrastructure is one of the most important determinants for success in data sharing (Gil-Garcia and Sayogo, 2016). Investments in high-quality IT infrastructure for data sharing, enabling effective data storage, data processing and data analysis are vital for successfully benefiting and, crucial for innovating from data (Krämer et al., 2020; BDVA, 2019). Particularly, organizations with little previous data sharing experience (or none at all) are deterred from data sharing due to the overwhelming complexity combined with high costs (Pauer et al., 2018). Small businesses are exceptionally challenged, as the extensive structural and financial efforts can create a threatening risk to the organization. Accordingly, the lack of an adequate technical infrastructure represents a major challenge for B2B data sharing.

### **3.4.2 Lack of Interoperability and Technical Standards**

In addition to inadequate technical infrastructure, insufficient data interoperability and missing data standards inhibit data sharing. Thereby, *data interoperability* is the ability to combine data sets from different sources (Gal and Rubinfeld, 2019) through compatible technical properties, enabling a seamless data exchange between multiple companies and digital sys-

tems (European Commission, 2017b). To ensure data interoperability, data standardization is required (European Commission, 2017b). *Data standardization* refers to specifying data attributes, structures and organization of data sets, data storage and data usage consistently along the data value chain (Gal and Rubinfeld, 2019). Across organizations, however, data sets often differ in terms of semantics, formats and storage systems (BDVA, 2017b; Arnaut et al., 2018b). The absence of a standard leads to technical obstacles, which inhibits data (re-)use and finally, results in a reduced willingness to share data (BDVA, 2017b; Gal and Rubinfeld, 2019). For a detailed discussion of the technological obstacles to the (re-)use of data, triggered by a lack of standards, see, e.g., Gal and Rubinfeld (2019).

While lacking standards inhibits sharing and using data, particularly when sharing partners involved do not pursue the same interests or offering competing products, widely-used standards are hard to achieve, as excluding said organizations and creating divergent standards is tempting (Martens et al., 2020). However, a lack of interoperability due to incompatible standards and insufficient standardization requires extensive additional efforts and investments to engage in and benefit from data sharing (BDVA, 2017a; European Commission, 2017b). Thus, technical barriers create potentially prohibitive costs and efforts to data sharing and thus inhibit doing so, even when legal and economic barriers are sufficiently resolved. In this context, the European Commission intends to put in place an legislative framework for data spaces, which facilitates data use, fosters data interoperability and prioritizes standardization activities (European Commission, 2020a).

## **4 Policy and Managerial Implications**

In analyzing the barriers to data sharing, this research offers two theoretical contributions and substantial practical implications, illustrated in Figure 2.

## 4.1 Contributions to Research

First, the existing literature on data sharing barriers from an economic perspective, was summarized and systematized offering a comprehensive overview and structure to the increasing number of studies. It is striking that both, academic and press articles repeatedly emphasize the tremendous value of data and data sharing, but very few empirical examples live up to these expectations. While extant literature shows an ambiguous stance on data sharing, repeatedly emphasizing its extensive benefits, while criticizing its lack in utilization (Otto, 2019; Jernigan et al., 2016; Koutroumpis et al., 2020). Aside from practitioner studies (see, e.g., Arnaut et al. (2018b), Barbero et al. (2018), Pauer et al. (2018)) and policy communications (see, e.g., European Commission (2020a, 2022)), academic research addresses legal barriers to data sharing, such as undefined data ownership (Koutroumpis et al., 2020; Richter and Slowinski, 2019; Kerber, 2016), lacking legal frameworks and regulatory interventions (Richter and Slowinski, 2019), and lacking technical support (Mineraud et al., 2016; Manyika et al., 2015; Gal and Rubinfeld, 2019), but barely economic hindrances. By summarizing and contrasting extant studies, this gap between expectations and realization of data sharing is explained, through intensively analyzing economic barriers in data sharing and their impact on strategic organizational challenges. This gap is addressed by developing a comprehensive explanation how data sharing is inhibited by the economic implication of the barriers for managerial decision-making.

While the areas of legal, economic and technical barriers may seem little surprising, however, the analysis reveals the severeness as well as the most important and prevalent empirical characteristics to the economic barriers. This systematic analysis and the comprehensive framework reveals why data sharing remains underutilized. Particularly this contributes to Elsaify and Hasan (2021), who explicitly emphasized a need for further research on data sharing among firms regarding competition and to the call by the European Commission for identifying and addressing barriers to corporate data sharing (European Commission, 2020b).

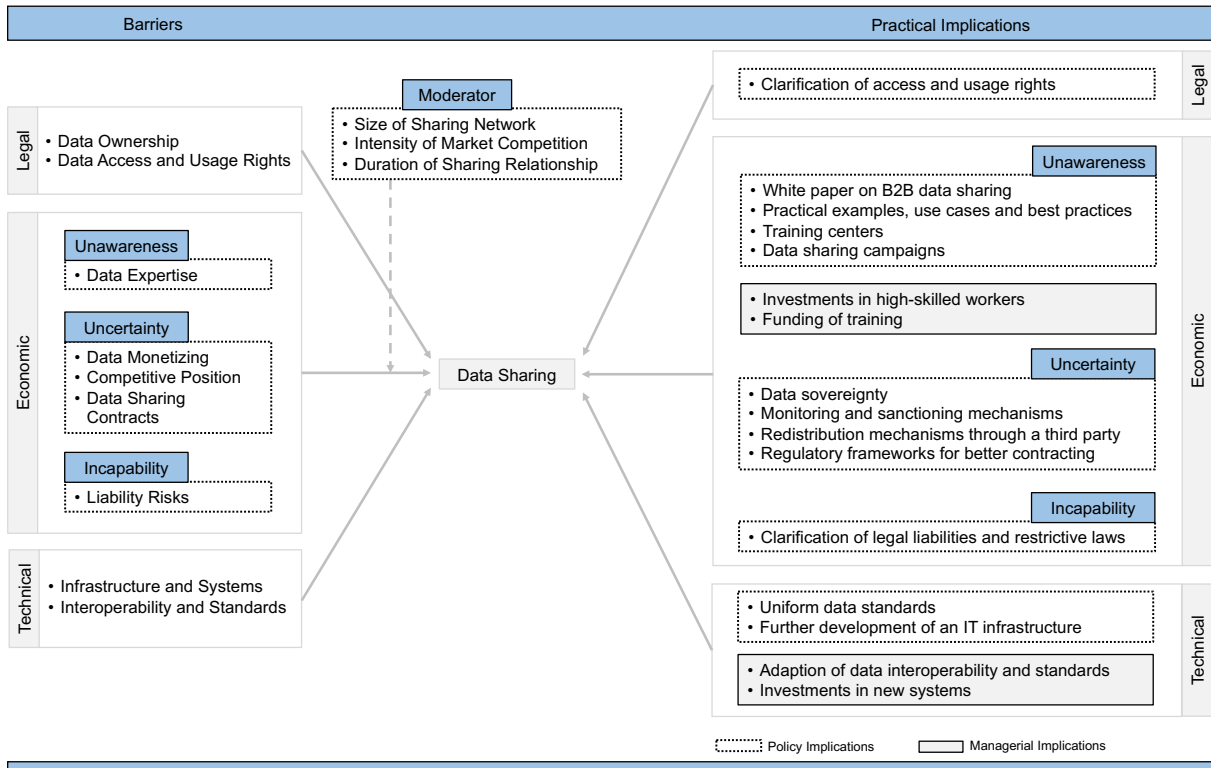


Figure 2: Barriers and Practical Implications.

Overcoming the legal and technical barriers and solving these issues requires efforts on a higher level than within sharing networks or between distinct organizations. Advancing through jurisdiction and regulations is a national or as in the case of the EU even a supra-national concern. Companies can not be expected to overcome these barriers on their own but need to rely on politically enabled change advancing data sharing. Several actions have already been taken by policymakers (European Commission, 2020a,b) and industry associations (BMW, 2020; International Data Spaces Association, 2022) to overcome these barriers. Despite the lively debate and the increase in regulations and initiatives, the potentials of data sharing still remain underutilized. Thus, this paper takes a different stance and reflects barriers to data sharing in the light of economic implications. From this perspective, new impulses to the ongoing debate are contributed.

Second, by linking the economic barriers to strategic problems arising in organizations that consider data sharing, the barriers are linked to organizational decision-making ex-



plaining why organizations are reluctant to or incapable of sharing data. This allows a prioritization of issues inhibiting data sharing and thereby, the findings contribute to an understanding how B2B data sharing could be improved. While unawareness of data sharing and subsequent opportunities is challenging, it can be widely solved within an organization by diverting additional resources in form of personnel, time or funding to exploring the topic of data sharing. Solving the uncertainty preventing extensive data sharing, can widely be accomplished within data sharing networks and thus within inter-organizational relationships such as industrial, business and trade associations. The strategic problem of incapability, however, is more challenging and requires for example additional or extensively revised laws, explicitly directed at data sharing.

## **4.2 Policy Implications**

The findings provide insightful recommendations for policymakers (see Figure 2). While pointing out the barriers to data sharing in this paper, policymaker are necessary to address and overcome these in all three areas - legal, economic and technical - effectively to enable data sharing.

**Legal** A major problem is the lack of property rights for non-personal, machine-generated data, causing uncertainty in data sharing. However, as mentioned in the findings, it has been found that an ownership right for data would cause more problems than benefits. Therefore, more attention to data access and usage rights is needed and essentially this topic requires additional clarification to advance managerial decision-making on favor of data sharing. To access and share their data, organizations depend on enforceable laws enabling data access.

Additionally, B2B data sharing could benefit from enforceable usage rights. Organizations lack lawfully binding restrictions to data processing and data usage that allow sharing partners to prevent or punish data misappropriation. The lack of usage rights is a key problem that triggers incapability in data sharing as managerial decision-making is strictly limited

in the light of extensive economic risks entailed. Thus, this issues needs to be addressed more extensively both academically and politically.

**Economic Unawareness** Companies and decision-makers lament the lack of easily accessible information on how data can be shared and used effectively and what to do in the case of misuse of shared data. Providing practitioner oriented, concise guidelines could additionally help navigating the complexity of data sharing and counteract the economic unawareness. Upon establishing a regulatory framework, a white paper on B2B data sharing could offer additional clarity and best practice advice. Such guidance can help companies assess data sharing opportunities, clarify economic, legal and political expectation and ultimately encourage data sharing investments. Practical examples, use cases and best practices might further help to understand the concept of data sharing and provide a clear path to follow. To this end, data courses at adult education centers could be offered, additional training centers established, and data sharing campaigns launched where employees can receive further training and regularly refresh their data knowledge.

**Economic Uncertainty** In terms of economic uncertainty, in particular the problem of usage control must be solved. Herein, an ecosystem that provides an appropriate form of data sovereignty, enabling individuals and organizations to determine and control how their data is used (Jarke et al., 2019), is key (Munoz-Arcentales et al., 2019; Otto, 2019). Data sovereignty contributes to a solution for confidentiality and security requirements in the area of non-personal data (BDVA, 2019). However, this requires clear guidelines on how to implement data sovereignty in dynamic business ecosystems which do not yet exist (BDVA, 2019).

To reduce the concerns about poor data quality, additional monitoring and sanctioning mechanisms for B2B data sharing networks are crucial (SINE, 2020), especially if no long-term sharing relationship exists and thus, trust could not yet be established. Regarding uncertainty in data monetizing, a business model and redistribution mechanism through

a third party could help to reduce this hurdle (Martens et al., 2020). In this vein, the European Commission has already taken measures to facilitate data sharing and to increase data availability through trusted data intermediaries, anchored in the Data Governance Act (DGA, 2020) (European Commission, 2020b). However, this entails high compliance costs for intermediaries and major uncertainties for companies, particularly with regard to a) measures to protect disclosure of non-personal data, b) the GDPR and antitrust law, and c) responsibilities of regulatory authorities. Yet, there is a registration requirement for potential intermediaries, but as long as there is no individual verification or requirement for appropriate certification, it will continue to be difficult for companies to build trust and uncertainties will remain.

Incomplete contracts are problematic in their own right (Grossman and Hart, 1986; Hart, 1988), but lacking definitions of data related responsibilities, liabilities and unclear designated conflict management institution and processes dramatically reinforce the behavioral uncertainties. The European Commission has addressed this issue by specifying key principles for non-personal, machine-generated data sharing contracts (European Commission, 2018a,b). However, as long as there are no enforceable regulatory frameworks established, contracting will not comprehensively improve data sharing.

**Economic Incapability** Companies fear the inability to enforce violations in data sharing agreements and thereby resulting extensive liability costs; making them economically incapable of sharing data. The European Commission intends to clarify the legal liability for data use with the recently published Data Act (European Commission, 2020a, 2022). However, to what extent this will take effect remains to be seen.

Another relevant topic are prohibitively restrictive legislation, such as competition law, rendering organization (perceivably) incapable of sharing data. Firms are restricted in deciding whether, which and to what extent they share data with competitors. While of course the importance of competition law is acknowledged, the current lack of definitions easily

places data sharing under general suspicion of competitive distortion. Thus, the EU must create a distinct legal framework for data sharing, including distinct and binding definitions for data, data types as well as corresponding rights and liabilities. Therein, clarification when restrictive laws, such as competition law, take effect, needs to be provided. For different data, the likeliness of abuse varies and strict data usage agreements could allow sharing with limited partners meeting predefined criteria.

**Technical** The technical hurdles should be addressed by policymakers and industry representatives. Politics must set up a legislative framework for interoperability standards that enhance the technical feasibility of data sharing. The industry in return, needs to build a conforming data sharing infrastructure. The European Commission initiatives for shared European data spaces to facilitate data use, encourage data standardization activities and foster data interoperability is welcomed (European Commission, 2020a). Further, *GAIA-X*, an initiative of European representatives, is aiming to create a trusted and sovereign data infrastructure for Europe to ensure secure and reliable data access (BMW, 2020). However, uniform standards still need to be established, and IT infrastructure needs to be further deployed and expanded to effectively mitigate this technical barrier.

### **4.3 Managerial Implications**

In addition to political measures, companies themselves must invest to successfully participate in data sharing (see Figure 2). In this context, economic unawareness and technical problems are barriers that companies need to address.

**Economic Unawareness** First of all, managers need to understand the benefits of data sharing and how to take advantage of it. Data science capabilities help to determine the value of data and understand how to leverage data from other organizations (Wu et al., 2020; Tambe, 2014), leading to more data sharing among sharing partners (Elsaify and

Hasan, 2021). Thus, managers should increasingly invest in high-skilled workers and fund employee training to build data expertise and keep up with technological progress.

**Technical** To build technical capabilities for sharing data, technical tools and projects that promote the adoption of data interoperability and standards need further support and funding. Involved sharing partners and platforms need to agree, e.g., on data formats, and standards for data sharing application programming interfaces (APIs) (EBA Open Banking Working Group, 2018). However, standards do not foster data sharing per se, but create the necessary technological infrastructure for the development and dissemination of data and thus reinforcing incentives for sharing and for addressing technical barriers (Gal and Rubinfeld, 2019). For example, the International Data Spaces Association (IDSA), a worldwide alliance of multiple stakeholder organizations has been developed a standardized reference architecture model - the *International Data Space (IDS)*<sup>10</sup> - for secure and trustworthy data sharing via platforms (Otto and Jarke, 2019; Jarke et al., 2019). Nevertheless, companies should invest additionally to adapt their systems to new structures in order to create the technically necessary conditions for data sharing.

## 5 Conclusions and Directions for Future Research

Data is an essential resource for businesses and society and a key driver for innovation. To leverage the full potential of industrial data, data sharing between businesses is crucial. However, B2B data sharing of non-personal, machine-generated data is not yet well-established. The developed framework explains the underutilization of data sharing as a result of existing barriers and the strategic challenges these pose for organizations. By shedding light on the economic implications of data sharing and identifying and classifying barriers to data sharing, understanding why firms are still reluctant or simply unable to share data, is enhanced.

---

<sup>10</sup><https://internationaldataspaces.org>

Despite the careful research process, some limitations of the proposed framework have to be acknowledged that may prompt future research. First, uncertainty in this study mainly emphasizes behavioural uncertainty. Further studies could integrate environmental uncertainty related to data sharing. Particularly as technology rapidly advances, data sharing will evolve in ways that are hard to predict. This may impact efforts for engaging in data sharing. Second, it should be noted that even if all technical and regulatory requirements are met, trustworthy data sharing remains challenging (BDVA, 2019), particularly in horizontal relationships. While trust building is measured by the length of the relationship in this study, further research should examine specific trust-building measures in data sharing. Lastly, based on this overview, further detailed (empirical) research can consider the barriers at a finer-grained level, especially when it comes to implementing effective practices for data sharing in organizations. Such a more detailed exploration may yield potential interventions to mitigate remaining barriers. Intervention and practices will be critical in determining appropriate strategies to promote data sharing to organizations.

## References

- Arnaut, C., M. Pont, E. Scaria, A. Berghmans, and S. Leconte (2018a). Study on data sharing between companies in europe - case studies. <https://op.europa.eu/en/publication-detail/-/publication/004fcf02-49c7-11e8-be1d-01aa75ed71a1/language-en/format-PDF/source-search>. Accessed September 8, 2021.
- Arnaut, C., M. Pont, E. Scaria, A. Berghmans, and S. Leconte (2018b). Study on data sharing between companies in europe. final report. <https://op.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en>. Accessed August 18, 2021.
- Arrow, K. J. (1962). Economic welfare and the allocation of resources for invention. In *The Rate and Direction of Inventive Activity: Economic and Social Factors*, pp. 609–626.

Princeton: Princeton University Press.

Asbroeck, B. V., J. Debussche, and J. César (2017). Building the european data economy. data ownership. [https://sites-twobirds.vuture.net/1/773/uploads/white-paper-ownership-of-data-\(final\).PDF](https://sites-twobirds.vuture.net/1/773/uploads/white-paper-ownership-of-data-(final).PDF). Accessed May 11, 2021.

Barbero, M., D. Cocoru, H. Graux, A. Hillebrand, F. Linz, D. Osimo, A. Siede, and P. Wauters (2018). Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability. <https://op.europa.eu/en/publication-detail/-/publication/74cca30c-4833-11e8-be1d-01aa75ed71a1/language-en>. Accessed August 13, 2021.

BDVA (2017a). Bdva’s response to the european commission’s communication and consultation ”building a european data economy”. [https://www.bdva.eu/sites/default/files/BDVA%20position%20to%20Buidling%20a%20European%20Data%20Economy%20Consultation\\_final.pdf](https://www.bdva.eu/sites/default/files/BDVA%20position%20to%20Buidling%20a%20European%20Data%20Economy%20Consultation_final.pdf). Big Data Value Association. Position Paper.

BDVA (2017b). Big data value sria 3.0. [https://www.bdva.eu/sites/default/files/EuropeanBigDataValuePartnership\\_SRIA\\_\\_v3.pdf](https://www.bdva.eu/sites/default/files/EuropeanBigDataValuePartnership_SRIA__v3.pdf). Big Data Value Association. Position Paper.

BDVA (2019). Towards a european data sharing space. enabling data exchange and unlocking ai potential. [https://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpace%20PositionPaper\\_April2019\\_V1.pdf](https://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpace%20PositionPaper_April2019_V1.pdf). Big Data Value Association. Position Paper.

BMWi, F. (2020). Gaia-x: A federated data infrastructure for europe. <https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html>. Accessed January 11, 2022.

Botelho, T. L. (2018). Here’s an opportunity: Knowledge sharing among competitors as a response to buy-in uncertainty. *Organization Science* 29(6), 1033–1055.

- Bouchbout, K. and Z. Alimazighi (2008). A framework for identifying the critical factors affecting the decision to adopt and use inter-organizational information systems. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering* 2(7), 723–730.
- Butler Jr, J. K. (1999). Trust expectations, information sharing, climate of trust, and negotiation effectiveness and efficiency. *Group & Organization Management* 24(2), 217–238.
- Cammarota, G., G. Ianiro, A. Ahern, C. Carbone, A. Temko, M. J. Claesson, A. Gasbarrini, and G. Tortora (2020). Gut microbiome, big data and machine learning to promote precision medicine for cancer. *Nature Reviews Gastroenterology & Hepatology* 17(10), 635–648.
- Cao, Q. H., M. Giyyarpuram, R. Farahbakhsh, and N. Crespi (2020). Policy-based usage control for a trustworthy data sharing platform in smart cities. *Future Generation Computer Systems* 107(C), 998–1010.
- Caricom Competition Commission (2021). A note on information sharing among competing firms. [http://www.caricomcompetitioncommission.com/images/pdf/info\\_sharing\\_rev.pdf](http://www.caricomcompetitioncommission.com/images/pdf/info_sharing_rev.pdf). Accessed July 21, 2021.
- Cazier, J. A., B. Shao, and R. D. S. Louis (2007). Sharing information and building trust through value congruence. *Information Systems Frontiers* 9(5), 515–529.
- Chen, M., S. Mao, and Y. Liu (2014). Big data: A survey. *Mobile Networks and Applications* 19(2), 171–209.
- Childerhouse, P., R. Hermiz, R. Mason-Jones, A. Popp, and D. R. Towill (2003). Information flow in automotive supply chains—identifying and learning to overcome barriers to change. *Industrial Management & Data Systems* 103(7), 491–502.



- Clemons, E. K. and L. M. Hitt (2004). Poaching and the misappropriation of information: Transaction risks of information exchange. *Journal of Management Information Systems* 21(2), 87–107.
- Cruijssen, F., W. Dullaert, and H. Fleuren (2007). Horizontal cooperation in transport and logistics: a literature review. *Transportation Journal* 46(3), 22–39.
- Das, T. K. and B.-S. Teng (1998). Between trust and control: Developing confidence in partner cooperation in alliances. *Academy of Management Review* 23(3), 491–512.
- Dawes, S. S. (1996). Interagency information sharing: Expected benefits, manageable risks. *Journal of Policy Analysis and Management* 15(3), 377–394.
- De Prieëlle, F., M. De Reuver, and J. Rezaei (2020). The role of ecosystem data governance in adoption of data platforms by internet-of-things data providers: Case of dutch horticulture industry. *IEEE Transactions on Engineering Management*, 1–11.
- Drexler, J. (2018). Data access and control in the era of connected devices. study on behalf of the european consumer organisation beuc. [https://www.beuc.eu/publications/beuc-x-2018-121\\_data\\_access\\_and\\_control\\_in\\_the\\_area\\_of\\_connected\\_devices.pdf](https://www.beuc.eu/publications/beuc-x-2018-121_data_access_and_control_in_the_area_of_connected_devices.pdf). Accessed August 11, 2021.
- Duch-Brown, N., B. Martens, and F. Mueller-Langer (2017). The economics of ownership, access and trade in digital data. <https://ec.europa.eu/jrc/sites/default/files/jrc104756.pdf>. JRC Technical Reports. Working Paper.
- EBA Open Banking Working Group (2018). B2b data sharing: Digital consent management as a driver for data opportunities. [https://eba-cms-prod.azurewebsites.net/media/azure/production/1815/eba\\_2018\\_obwg\\_b2b\\_data\\_sharing.pdf](https://eba-cms-prod.azurewebsites.net/media/azure/production/1815/eba_2018_obwg_b2b_data_sharing.pdf). Accessed April 18, 2021.
- Elsaify, M. and S. Hasan (2021). Data exchanges among firms. *Digital Business* 1(2), 1–11.

European Commission (2011). Digital agenda: Turning government data into gold. [https://ec.europa.eu/commission/presscorner/detail/en/IP\\_11\\_1524](https://ec.europa.eu/commission/presscorner/detail/en/IP_11_1524). European Commission Press Release.

European Commission (2014). Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions "towards a thriving data-driven economy". Communication COM(2014) 442 final, European Commission, Brussels.

European Commission (2016). Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions "digitising european industry reaping the full benefits of a digital single market". Communication COM(2016) 180 final, European Commission, Brussels.

European Commission (2017a). Commission staff working document on the free flow of data and emerging issues of the european data economy "building a european data economy". Communication SWD(2017) 2 final, European Commission, Brussels.

European Commission (2017b). Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions "building a european data economy". Communication COM(2017) 9 final, European Commission, Brussels.

European Commission (2018a). Commission staff working document. guidance on sharing private sector data in the european data economy. Communication SWD(2018) 125 final, European Commission, Brussels.

European Commission (2018b). Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions "towards a common european data space". Communication COM(2018) 232 final, European Commission, Brussels.

European Commission (2019). Sme panel consultation on b2b data-sharing principles and guidance. report on the results. <https://digital-strategy.ec.europa.eu/en/library/sme-panel-consultation-b2b-data-sharing>. European Commission Report.

European Commission (2020a). Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions "a european strategy for data". Communication COM(2020) 66 final, European Commission, Brussels.

European Commission (2020b). Regulation of the european parliament and of the council on european data governance (data governance act). Communication COM(2020) 767 final, European Commission, Brussels.

European Commission (2022). Regulation of the european parliament and of the council on harmonised rules on fair access to and use of data (data act). Communication COM(2022) 68 final, European Commission, Brussels.

European Union (2014). Regulation (eu) no 536/2014 of the european parliament and of the council of 16 april 2014 on clinical trials on medicinal products for human use, and repealing directive 2001/20/ec. [https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/reg\\_2014\\_536/reg\\_2014\\_536\\_en.pdf](https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/reg_2014_536/reg_2014_536_en.pdf). Accessed April 19, 2021.

European Union (2015). Directive (eu) 2015/2366 of the european parliament and of the council of 25 november 2015 on payment services in the internal market, amending directives 2002/65/ec, 2009/110/ec and 2013/36/eu and regulation (eu) no 1093/2010, and repealing directive 2007/64/ec. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32015L2366&from=DE>. Accessed April 19, 2021.

European Union (2018). Regulation (eu) 2018/1807 of the european parliament and of the council of 14 november 2018 on a framework for the free flow of non-personal data in

the european union. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018R1807&from=EN>. Accessed April 19, 2021.

European Union (2019). Directive (eu) 2019/1024 of the european parliament and of the council of 20 june 2019 on open data and the re-use of public sector information. [http://publications.europa.eu/resource/cellar/a6ef4c41-97eb-11e9-9369-01aa75ed71a1.0006.03/DOC\\_1](http://publications.europa.eu/resource/cellar/a6ef4c41-97eb-11e9-9369-01aa75ed71a1.0006.03/DOC_1). Accessed April 19, 2021.

Fawcett, S. E., P. Osterhaus, G. M. Magnan, J. C. Brau, and M. W. McCarter (2007). Information sharing and supply chain performance: the role of connectivity and willingness. *Supply Chain Management* 12(5), 358–368.

Federation of German Industries (BDI) (2020). Public consultation procedure on the eu data strategy. [https://english.bdi.eu/media/themenfelder/digitalisierung/publikationen/20200529\\_BDI\\_Statement\\_EU-Data\\_Strategy\\_EN.pdf](https://english.bdi.eu/media/themenfelder/digitalisierung/publikationen/20200529_BDI_Statement_EU-Data_Strategy_EN.pdf). BDI statement on the EU data strategy.

Gal, M. S. and D. L. Rubinfeld (2019). Data standardization. *New York University Law Review* 94(4), 737–770.

Gal-Or, E. and A. Ghose (2005). The economic incentives for sharing security information. *Information Systems Research* 16(2), 186–208.

GDPR.EU (2021a). Art. 4 gdpr, definitions. <https://gdpr.eu/article-4-definitions>. Accessed July 26, 2021.

GDPR.EU (2021b). Art. 82 gdpr, definitions. <https://gdpr.eu/article-82-data-subjects-right-to-compensation-and-liability/>. Accessed July 26, 2021.

- GDPR.EU (2021c). Art. 83 gdpr, definitions. <https://gdpr.eu/article-83-conditions-for-imposing-administrative-fines/>. Accessed July 26, 2021.
- GDPR.EU (2021d). Art. 84 gdpr, definitions. <https://gdpr.eu/article-84-member-state-penalties/>. Accessed July 26, 2021.
- Gil-García, J. R. and T. A. Pardo (2005). E-government success factors: Mapping practical tools to theoretical foundations. *Government Information Quarterly* 22(2), 187–216.
- Gil-García, J. R. and D. S. Sayogo (2016). Government inter-organizational information sharing initiatives: Understanding the main determinants of success. *Government Information Quarterly* 33(3), 572–582.
- Green, H. (2016). The internet of things in the cognitive era: Realizing the future and full potential of connected devices. <https://www.ibm.com/common/ssi/cgi-bin/ssialias?%20htmlfid=WWW12366USEN>. IBM Watson IoT. Whitepaper.
- Grossman, S. J. and O. D. Hart (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy* 94(4), 691–719.
- Hall, D. W. (2020). Growing the artificial intelligence industry in the uk. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/652097/Growing\\_the\\_artificial\\_intelligence\\_industry\\_in\\_the\\_UK.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf). Accessed April 19, 2021.
- Hart, O. D. (1988). Incomplete contracts and the theory of the firm. *Journal of Law, Economics, and Organization* 4(1), 119–139.
- HERE (2021a). All your location data in one platform. <https://www.here.com>. Accessed July 30, 2021.

- HERE (2021b). Driving the automotive industry forward with location technology. <https://www.here.com/solutions/automotive>. Accessed July 30, 2021.
- HERE (2021c). Here marketplace. <https://www.here.com/platform/marketplace>. Accessed July 30, 2021.
- Hsu, M.-H. and C.-M. Chang (2014). Examining interpersonal trust as a facilitator and uncertainty as an inhibitor of intra-organisational knowledge sharing. *Information Systems Journal* 24(2), 119–142.
- International Data Spaces Association (2022). International data spaces. the future of the data economy is here. <https://internationaldataspaces.org>. Accessed January 10, 2022.
- Isaac, R. M., J. M. Walker, and S. H. Thomas (1984). Divergent evidence on free riding: An experimental examination of possible explanations. *Public Choice* 43(2), 113–149.
- Janssen, M., Y. Charalabidis, and A. Zuiderwijk (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management* 29(4), 258–268.
- Jarke, M., B. Otto, and S. Ram (2019). Data sovereignty and data space ecosystems. *Business & Information Systems Engineering* 61(5), 549–550.
- Jentzsch, N., G. Sapi, and I. Suleymanova (2013). Targeted pricing and customer data sharing among rivals. *International Journal of Industrial Organization* 31(2), 131–144.
- Jernigan, S., S. Ransbotham, and D. Kiron (2016). Data sharing and analytics drive success with iot-creating business value with the internet of things. *MIT Sloan Management Review* 58(1), 1–17.
- Jin, Y., H. C. B. Lee, S. Ba, and J. Stallaert (2021). Winning by learning? effect of knowledge sharing in crowdsourcing contests. *Information Systems Research* 32(3), 836–859.

- Jones, C. I. and C. Tonetti (2020). Nonrivalry and the economics of data. *American Economic Review* 110(9), 2819–2858.
- Kembro, J., K. Selviaridis, and D. Näslund (2014). Theoretical perspectives on information sharing in supply chains: a systematic literature review and conceptual framework. *Supply Chain Management* 19(5/6), 609–625.
- Kerber, W. (2016). A new (intellectual) property right for non-personal data? an economic analysis. <https://www.econstor.eu/bitstream/10419/155649/1/870294326.pdf>. MAGKS Joint Discussion Paper Series in Economics, No. 37-2016. Working Paper.
- Kerr, D., D. Arnold, J.-Y. Blay, C. Buske, A. Carrato, W. Gerritsen, and M. Peeters (2020). The oncology data network (odn): A collaborative european data-sharing platform to inform cancer care. *The Oncologist* 25(1), e1–e4.
- Kimmel, M. J., D. G. Pruitt, J. M. Magenau, E. Konar-Goldband, and P. J. Carnevale (1980). Effects of trust, aspiration, and gender on negotiation tactics. *Journal of Personality and Social Psychology* 38(1), 9–22.
- Klein, R. and A. Rai (2009). Interfirm strategic information flows in logistics supply chain relationships. *Mis Quarterly* 33(4), 735–762.
- Kollock, P. (1994). The emergence of exchange structures: An experimental study of uncertainty, commitment, and trust. *American Journal of Sociology* 100(2), 313–345.
- Koutroumpis, P., A. Leiponen, and L. D. Thomas (2020). Markets for data. *Industrial and Corporate Change* 29(3), 645–660.
- Krämer, J., D. Schnurr, and S. B. Micova (2020). The role of data for digital markets contestability: case studies and data access remedies. [https://cerre.eu/wp-content/uploads/2020/08/cerre-the\\_role\\_of\\_data\\_for\\_digital\\_markets\\_contestability\\_](https://cerre.eu/wp-content/uploads/2020/08/cerre-the_role_of_data_for_digital_markets_contestability_)

- case\_studies\_and\_data\_access\_remedies-september2020.pdf. Centre on Regulation in Europe (CERRE) Policy Report.
- Kumar, R. S. and S. Pugazhendhi (2012). Information sharing in supply chains: An overview. *Procedia Engineering* 38, 2147–2154.
- Lado, A. A., N. G. Boyd, and S. C. Hanlon (1997). Competition, cooperation, and the search for economic rents: A syncretic model. *Academy of Management Review* 22(1), 110–141.
- Lee, H. L. and S. Whang (2000). Information sharing in a supply chain. *International Journal of Manufacturing Technology and Management* 1(1), 79–93.
- Li, L. and H. Zhang (2008). Confidentiality and information sharing in supply chain coordination. *Management Science* 54(8), 1467–1481.
- Li, S. and B. Lin (2006). Accessing information sharing and information quality in supply chain management. *Decision Support Systems* 42(3), 1641–1656.
- Lippert, S. K. and C. Govindarajulu (2006). Technological, organizational, and environmental antecedents to web services adoption. *Communications of the IIMA* 6(1), 146–158.
- Liu, Q. and K. Serfes (2006). Customer information sharing among rival firms. *European Economic Review* 50(6), 1571–1600.
- Madlberger, M. (2008). Interorganizational collaboration in supply chain management: what drives firms to share information with their trading partners? In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*, pp. 1–10. IEEE.
- Mäkipää, M. (2006). Inter-organizational information systems in cooperative inter-organizational relationships: Study of the factors influencing to success. In R. Suomi, R. Cabral, J. F. Hampe, A. Heiklula, J. Jirveliiinen, and E. Kosluvaara (Eds.), *IFIP*



- International Federation for Information Processing, Project E-Society: Building Bricks*, Volume 226, pp. 68–81. Boston: Springer.
- Manyika, J., M. Chui, P. Bisson, J. Woetzel, R. Dobbs, J. Bughin, and D. Aharon (2015). [https://www.mckinsey.com/~media/McKinsey/Industries/Technology%20Media%20and%20Telecommunications/High%20Tech/Our%20Insights/The%20Internet%20of%20Things%20The%20value%20of%20digitizing%20the%20physical%20world/Unlocking\\_the\\_potential\\_of\\_the\\_Internet\\_of\\_Things\\_Executive\\_summary.pdf](https://www.mckinsey.com/~media/McKinsey/Industries/Technology%20Media%20and%20Telecommunications/High%20Tech/Our%20Insights/The%20Internet%20of%20Things%20The%20value%20of%20digitizing%20the%20physical%20world/Unlocking_the_potential_of_the_Internet_of_Things_Executive_summary.pdf). Accessed October 8, 2021.
- Martens, B., A. de Streel, I. Graef, T. Tombal, and N. Duch-Brown (2020). Business-to-business data sharing: An economic and legal analysis. <https://www.econstor.eu/bitstream/10419/227701/1/1726726282.pdf>. JRC Technical Report. Working Paper.
- Martens, B. and N. Duch-Brown (2020). The economics of business-to-government data sharing. <https://ec.europa.eu/jrc/sites/default/files/jrc119947.pdf>. JRC Technical Report. Working Paper.
- Mattioli, M. (2014). Disclosing big data. *Minnesota Law Review* 99(2), 535–583.
- McLoughlin, B. (2021). This is what the dashboard of the future looks like. <https://360.here.com/unity-3d-dashboards>. Accessed July 30, 2021.
- Mellewigt, T., A. Madhok, and A. Weibel (2007). Trust and formal contracts in inter-organizational relationships—substitutes and complements. *Managerial and Decision Economics* 28(8), 833–847.
- Michener, W. K. (2015). Ecological data sharing. *Ecological Informatics* 29(1), 33–44.
- Mineraud, J., O. Mazhelis, X. Su, and S. Tarkoma (2016). A gap analysis of internet-of-things platforms. *Computer Communications* 89(C), 5–16.

Ministry of Economic Affairs and Climate Policy (2019). Dutch digitalisation strategy: Dutch vision on data sharing between businesses. <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&ved=2ahUKEwj2kd0qmc3jAhVFD0wKHdieBb4QFjADegQIBBAC&url=https%3A%2F%2Fwww.government.nl%2Fbinaries%2Fgovernment%2Fdocuments%2Freports%2F2019%2F02%2F01%2Fdutch-vision-on-data-sharing-between-businesses%2FDutch%2Bvision%2Bon%2Bdata%2Bsharing%2Bbetween%2Bbusinesses.pdf&usq=A0vVaw10VaXErN1EgRhDZD8h19Gi>. Accessed May 26, 2021.

Munoz-Arcentales, A., S. López-Pernas, A. Pozo, Á. Alonso, J. Salvachúa, and G. Huecas (2019). An architecture for providing data usage and access control in data sharing ecosystems. *Procedia Computer Science* 160, 590–597.

Natwichai, J., X. Li, and M. Orłowska (2005). Hiding classification rules for data sharing with privacy preservation. In A. Tjoa and J. Trujillo (Eds.), *Data Warehousing and Knowledge Discovery*, Volume 3589 of *Lecture Notes in Computer Science*, pp. 468–477. Berlin, Heidelberg: Springer.

New York Law School (2012). Horizontal information exchanges. <https://www.eucomplaw.com/article-101/horizontal-information-exchange/>. Accessed April 9, 2021.

Ngiam, K. Y. and W. Khor (2019). Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology* 20(5), 262–273.

Norman, P. M. (2001). Are your secrets safe? knowledge protection in strategic alliances. *Business Horizons* 44(6), 51–61.

OECD (2015). Data-driven innovation: big data for growth and well-being. <http://dx.doi.org/10.1787/9789264229358-en>. Accessed May 24, 2021.

Otto, B. (2019). Interview with reinhold achatz on “data sovereignty and data ecosystems”. *Business & Information Systems Engineering* 61(5), 635–636.

- Otto, B. and M. Jarke (2019). Designing a multi-sided data platform: findings from the international data spaces case. *Electronic Markets* 29(4), 561–580.
- Özer, Ö., Y. Zheng, and K.-Y. Chen (2011). Trust in forecast information sharing. *Management Science* 57(6), 1111–1137.
- Pala, A. and J. Zhuang (2019). Information sharing in cybersecurity: A review. *Decision Analysis* 16(3), 172–196.
- Pauer, A., L. Nagel, T. Fedkenhauser, Fritzsche-Sterr, and Resetko (2018). Data exchange as a first step towards data economy. <https://www.pwc.de/en/digitale-transformation/data-exchange-as-a-first-step-towards-data-economy.pdf>. Accessed April 3, 2021.
- Perkmann, M. and H. Schildt (2015). Open data partnerships between firms and universities: The role of boundary organizations. *Research Policy* 44(5), 1133–1143.
- Praditya, D., M. Janssen, and R. Sulastris (2017). Determinants of business-to-government information sharing arrangements. *Electronic Journal of e-Government* 15(1), 44–55.
- Pretschner, A., M. Hilty, D. Basin, C. Schaefer, and T. Walter (2008). Mechanisms for usage control. In *Proceeding ASIACCS '08 Proceedings of the 2008 ACM symposium on Information, computer and communications security*, pp. 240–244. Association for Computing Machinery (ACM).
- Pretschner, A., E. Lovat, and M. Büchler (2011). Representation-independent data usage control. In J. Garcia-Alfaro, G. Navarro-Arribas, N. Cuppens-Boulahia, and S. de Capitani di Vimercati (Eds.), *Data Privacy Management and Autonomous Spontaneous Security*, pp. 122–140. Berlin, Heidelberg: Springer.

- Ramon Gil-Garcia, J., I. Chengalur-Smith, and P. Duchessi (2007). Collaborative e-government: impediments and benefits of information-sharing projects in the public sector. *European Journal of Information Systems* 16(2), 121–133.
- Richter, H. and P. R. Slowinski (2019). The data sharing economy: on the emergence of new intermediaries. *IIC - International Review of Intellectual Property and Competition Law* 50(1), 4–29.
- Rindfleisch, A. (2000). Organizational trust and interfirm cooperation: an examination of horizontal versus vertical alliances. *Marketing Letters* 11(1), 81–95.
- Rousseau, D. M., S. B. Sitkin, R. S. Burt, and C. Camerer (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review* 23(3), 393–404.
- Ruj, S., A. Nayak, and I. Stojmenovic (2011). Dacc: Distributed access control in clouds. In *2011 IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications*, pp. 91–98. IEEE.
- Rutella, S., M. A. Cannarile, S. Gnjatic, B. Gomes, J. Guinney, V. Karanikas, M. Karkada, J. M. Kirkwood, B. Kotlan, G. V. Masucci, et al. (2020). Society for immunotherapy of cancer clinical and biomarkers data sharing resource document: Volume i—conceptual challenges. *Journal for Immunotherapy of Cancer* 8(2), 1–8.
- Samaddar, S., S. Nargundkar, and M. Daley (2006). Inter-organizational information sharing: The role of supply network configuration and partner goal congruence. *European Journal of Operational Research* 174(2), 744–765.
- Samarati, P. and S. C. de Vimercati (2000). Access control: Policies, models, and mechanisms. In R. Focardi and R. Gorrieri (Eds.), *International School on Foundations of Security Analysis and Design*, pp. 137–196. Berlin, Heidelberg: Springer.

- Sandhu, R. S. and P. Samarati (1994). Access control: principle and practice. *IEEE Communications Magazine* 32(9), 40–48.
- Schweitzer, H. and M. Peitz (2017). Datenmärkte in der digitalisierten wirtschaft: Funktionsdefizite und regelungsbedarf? <https://ftp.zew.de/pub/zew-docs/dp/dp17043.pdf>. ZEW Discussion Paper.
- SINE (2020). How to make companies share truthful data. <https://sine.foundation/library/how-to-make-companies-share-truthful-data>. Accessed October 15, 2021.
- Snijders, C. and G. Keren (1999). Determinants of trust. In D. V. Budescu, I. Erev, and R. Zwick (Eds.), *Games and human behavior: Essays in honor of Amnon Rapoport*, pp. 355–385. Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers.
- Tambe, P. (2014). Big data investment, skills, and firm value. *Management Science* 60(6), 1452–1469.
- Tenopir, C., S. Allard, K. Douglass, A. U. Aydinoglu, L. Wu, E. Read, M. Manoff, and M. Frame (2011). Data sharing by scientists: practices and perceptions. *PloS ONE* 6(6), 1–21.
- Trkman, P. and K. C. Desouza (2012). Knowledge risks in organizational networks: An exploratory framework. *The Journal of Strategic Information Systems* 21(1), 1–17.
- Truong, H.-L., M. Comerio, F. De Paoli, G. Gangadharan, and S. Dustdar (2012). Data contracts for cloud-based data marketplaces. *International Journal of Computational Science and Engineering* 7(4), 280–295.
- Wiebe, A. (2017). Protection of industrial data – a new property right for the digital economy? *Journal of Intellectual Property Law & Practice* 12(1), 62–71.

Wixom, Barbara, H., I. M. Sebastian, and R. W. Gregory (2020). Data sharing 2.0: New data sharing, new value creation. [https://cisr.mit.edu/publication/2020\\_1001\\_DataSharing\\_WixomSebastianGregory](https://cisr.mit.edu/publication/2020_1001_DataSharing_WixomSebastianGregory). Accessed August 20, 2021.

Wu, L., L. Hitt, and B. Lou (2020). Data analytics, innovation, and firm productivity. *Management Science* 66(5), 2017–2039.

Zhao, X. and L. Xue (2012). Competitive target advertising and consumer data sharing. *Journal of Management Information Systems* 29(3), 189–222.

# Sharing needs Caring: Experimental Insights on the Optimal Design of B2B Data Sharing Platforms

## Authors

Jan Krämer, University of Passau, Germany.  
Nadine Stüdlein, University of Passau, Germany.  
Oliver Zierke, University of Passau, Germany.

## Presented at<sup>2</sup>

- Gesellschaft für experimentelle Wirtschaftsforschung e.V. (GfeW), 2019, Düsseldorf, Germany.
- International Conference on Wirtschaftsinformatik - Doctoral Consortium, 2020, Berlin, Germany.
- Colloquium for doctoral students, 2020, Passau, Germany.

## Abstract

*Data sharing between firms is widely believed to promote firms' ability to innovate and compete. To this end, Business-to-Business (B2B) data sharing platforms, such as 'GAIA-X', facilitate the sharing of data between firms. However, many of these initiatives struggle. Although firms find it advantageous to use the data of other firms, they are reluctant to share their own data because they fear that this data could be used to their competitive disadvantage. Robust economic insights on how to design B2B data sharing platforms in such an environment are scarce. We identify two important dimensions for the contractual and technological design architecture of B2B data sharing platforms: control and transparency. Control denotes the degree to which a firm can specify which other firm can access how much of its data. Transparency denotes a firm's information about other firm's sharing behavior and the resulting profits. Using an economic laboratory experiment, we find that both, control and transparency are important dimensions of the design of B2B data sharing platforms, as an increase in either of these dimensions can significantly increase a firm's propensity to share data with other (rival) firms. However, we also show that more control does not always lead to more data sharing, and that the impact of the control and transparency dimensions depend on the duration of the market relationships. Hence, our results bear important insights for success factors of B2B data sharing platforms.*

---

<sup>2</sup>The list does *not* include presentations held by co-authors.

# Sharing needs Caring: Experimental Insights on the Optimal Design of B2B Data Sharing Platforms

Jan Krämer\*      Nadine Stüdlein†      Oliver Zierke‡

## Abstract

Data sharing between firms is widely believed to promote firms' ability to innovate and compete. To this end, Business-to-Business (B2B) data sharing platforms, such as 'GAIA-X', facilitate the sharing of data between firms. However, many of these initiatives struggle. Although firms find it advantageous to use the data of other firms, they are reluctant to share their own data because they fear that this data could be used to their competitive disadvantage. Robust economic insights on how to design B2B data sharing platforms in such an environment are scarce. We identify two important dimensions for the contractual and technological design architecture of B2B data sharing platforms: control and transparency. Control denotes the degree to which a firm can specify which other firm can access how much of its data. Transparency denotes a firm's information about other firm's sharing behavior and the resulting profits. Using an economic laboratory experiment, we find that both, control and transparency are important dimensions of the design of B2B data sharing platforms, as an increase in either of these dimensions can significantly increase a firm's propensity to share data with other (rival) firms. However, we also show that more control does not always lead to more data sharing, and that the impact of the control and transparency dimensions depend on the duration of the market relationships. Hence, our results bear important insights for success factors of B2B data sharing platforms.

*Keywords:* Data sharing, B2B data markets, laboratory experiment

*JEL Codes:* C91, L10, L20

---

\*Corresponding author. Chair of Internet and Telecommunications Business, University of Passau, Dr.-Hans-Kapfnger-Str. 12, 94032 Passau, Germany. Email: jan.kraemer@uni-passau.de. Declarations of interest: none

†Chair of Internet and Telecommunications Business, University of Passau, Dr.-Hans-Kapfnger-Str. 12, 94032 Passau, Germany. Email: nadine.stuedlein@uni-passau.de. Declarations of interest: none

‡Chair of Internet and Telecommunications Business, University of Passau, Dr.-Hans-Kapfnger-Str. 12, 94032 Passau, Germany. Email: oliver.zierke@uni-passau.de. Declarations of interest: none



# 1 Introduction

Data emerges as the essential resource in the 21st century and is increasingly important to our economy and society. Data has become vital to many business models. It allows targeted and personalized advertisements in online retailing and is used to develop and enhance artificial intelligence, advancing and reshaping Internet of Things technologies decisively (Hall, 2020; Ministry of Economic Affairs and Climate Policy, 2019). In order to exploit the full potential of data, the accessibility of data is crucial (European Commission, 2020; Martens et al., 2020; De Prieëlle et al., 2020; Jarke et al., 2019). To this end, the European Commission is planning legislative measures to improve access to and use of data (European Commission, 2020). In particular, the Data Act, published by the European Commission in February 2022, is intended to support data sharing between firms. This is to be achieved by reviewing competition rules, clarifying the rights of use for jointly generated data and the legal liability of data use (European Commission, 2020). Furthermore, data initiatives are being developed to promote data sharing. For example, *GAIA-X*, an initiative of European representatives aiming to create a trusted and sovereign data infrastructure for Europe to ensure secure and reliable data access (BMW, 2020a).

However, although data sharing is essential in promoting firms' innovativeness and competitiveness (Ministry of Economic Affairs and Climate Policy, 2019; Pauer et al., 2018; Telekom, 2019), many firms are reluctant to share their own data, as it makes firms potentially vulnerable through a loss of control over internal data (Otto, 2019; Pauer et al., 2018; Norman, 2001) that could result in a reduction in their competitive edge (European Commission, 2020; Natwichai et al., 2005; Trkman and Desouza, 2012). Furthermore, the uncertainty about the investment and the data uses by other firms, as well as coordination problems from struggling in finding qualified sharing partners, hinder data sharing (Cabrera and Cabrera, 2002; European Commission, 2020; Loebbecke et al., 2016; Ministry of Economic Affairs and Climate Policy, 2019). Thus, many firms miss vital opportunities for their business activities and ultimately business performance.

In practice, there are already some data initiatives, which seek to facilitate structured sharing of data among firms. While large, international initiatives, such as *GAIA-X*, are still being developed, there are a number of smaller B2B data sharing platforms initiated by businesses. The *Data Intelligence Hub*, for example, is a marketplace for data and analysis tools by Deutsche Telekom, which allows firms to conclude individual data sharing contracts (Telekom, 2019). In other initiatives, such as the *Open Data Initiative* by Adobe, Microsoft, and SAP, firms share data on a collective cloud platform, that can be accessed by all participating firms (Microsoft, 2019).

Despite the varying contractual designs on who can share data with whom and under which conditions, on an abstract level the contractual designs of B2B data sharing platforms can be distinguished along two fundamental dimensions: control and transparency. *Control* is defined as the degree to which a firm can control other firms' data access, i.e., whether a firm can deny or individually limit other firms' access to its shared data. *Transparency* is the degree of information a firm has about other firms' sharing behavior and the resulting profits, i.e., whether a firm has full information about others sharing actions and profits or only about those in which it is actively involved.

Of the aforementioned initiatives, the *Data Intelligence Hub* gives companies full control over their data, as they can determine precisely which other firm can access their data, based on individual contracts. However, the degree of transparency is low, as each firm only knows the actions in which it is involved itself, and not those of the other companies. In the *Open Data Initiative*, on the other hand, firms no longer have control over their data once it is shared on the platform, but each firm knows exactly who has shared how much data, which gives them a high degree of transparency. *GAIA-X*, finally, which is a multi-purpose federated data sharing infrastructure, allows for various and fine-grained degrees of control and transparency that have to be defined depending on the context (BMW, 2020b).

From an economic perspective, data can be considered as a public good. Yet, there are important differences between a classic public good and data. In particular, data can

be contributed to the public domain without having to surrender it at the source. In this article, we investigate by means of an economic laboratory experiment how control and transparency on B2B data sharing platforms affect firms' willingness to share data. The experimental methodology allows us to systematically vary contract features for B2B data platforms along these two dimensions while holding all other environmental conditions, such as the number of B2B sharing partners and the value of data for the firms, constant. This would not be possible in the field. The impacts of the dimensions control and transparency on data sharing behavior is relevant for firms seeking to participate on various data sharing platforms, as well as for designing such data platforms, and for policymakers aiming to fostering B2B data sharing. Drawing on insights from previous experimental evidence, we conjecture that more control will have a positive effect on the amount of shared data, while ex-ante no clear effect of transparency on data sharing can be deduced. The results of our experiment highlight that firms' propensity to share data generally (weakly) increases with control, but that an intermediate level of control may already be sufficient. Furthermore, we find that a higher degree of transparency is conducive to more data sharing.

The remainder of this paper is structured as follows: In Section 2 we discuss related literature and previous experimental findings on the impact of control and transparency in other contexts. Next, in Section 3 we describe the underlying theoretical model and experimental design. Section 4 describes the experimental procedure and results are discussed in Section 5. Finally, we discuss our results in the context of the design of B2B data sharing platforms in Section 6 and conclude with limitations and directions for future research in Section 7.

## **2 Related Literature**

Our study is related to a wide range of the theoretical and the empirical literature, concerned with knowledge and information sharing within and between firms and the benefits of co-operation (see, e.g., Kim et al. (2005); Huang et al. (2007); Wang and Noe (2010); Bogers

(2011); Kembro et al. (2014); Lee and Whang (2000) for overviews). However, in all these scenarios, there is a dependency between the sharing partners working together in a firm, on a project or on a product, while in data sharing, every potential sharing partner can succeed even if its potential partners do not.

Some closer related theoretical models address firms sharing consumer data (Jentzsch et al., 2013; Krämer et al., 2019; Zhao and Xue, 2012; Ghosh et al., 2015). In these models firms share data only if their joint profit increases with data sharing and the profits are redistributed such that sharing partners are better off after data sharing. In reality, however, data sharing inherent bears a risk that data recipients exploit the data senders and do not redistribute the additional profits accordingly, leaving the data senders worse off. In such an arguably more realistic scenario where opportunistic behavior of firms is possible, from a theoretical perspective, no firm should share any data, leading to a collapse of sharing markets. Yet, in practice data sharing markets may still exist, because data senders and data recipients interact repeatedly and come to realize that by cooperating and reciprocal sharing of data, they can attain larger profits in the long run than by short run opportunistic behavior and no sharing. In our economic laboratory experiment, we mimic this strategic situation inherent to B2B data sharing markets, and test, which market environments (with respect to the dimensions control and transparency) are more conducive to elicit cooperation and data sharing by firms.

The strategic situation in B2B data sharing markets has close parallels to games that have been studied previously in behavioral economics. Specifically, similar to the strategic situation studied here, in public-good games, n-person prisoner dilemmas and oligopoly competition games, cooperation at the individual level is costly, since non-cooperation, while others cooperate, maximizes the individual gain. Further, in all these games, cooperation at the industry level is beneficial, since the gain for all entities is higher when all entities cooperate than when no one cooperates. However, it cannot be assumed that the insights from the literature discussed simply carry over to the data context. The strategic interaction

when sharing data differs in one important aspect: Due to the non-rivalry of data, sharing data (unlike sharing physical resources) does not directly reduce the profit of the sharing firm at the individual level, but rather generates indirect costs by making rival firms stronger and ensuing strategic risks. By contrast, in a prisoner's dilemma game, for example, a player directly receives a lower reward if he or she decides to cooperate. Similarly, in an oligopoly game, a firm directly reduces its profits when it behaves more collusively and raises its price above the market equilibrium price. Finally, in a standard public good game, a player immediately receives a lower reward from the private good when investing in the public good rather than the private good. In the case of data sharing, however, the profit of the firm is not directly affected by whether data was shared or not, as the data remains available to the firm after sharing. Instead, data sharing incurs only indirect costs through the (potential) opportunistic use of the data by the sharing partners.<sup>1</sup>

To the best of our knowledge, our study is the first economic laboratory experiment that considers cooperation in the context of data sharing, and thus, we cannot directly infer predictions of control and transparency from previous experiments. Instead, due to the aforementioned strategic similarities with public-good games, n-person prisoner dilemmas and oligopoly competition games we form expectations on the impact of these dimensions based on experimental studies of these games, noting that there are important strategic differences between them and our context.

**Control:** There is no exact replication of control over data sharing in other experiments. However, control in the broader sense, i.e., control over the relationships to others, has been investigated in many experiments. One of the most common ways of investigating more control in public good games, n-person prisoner's dilemma games and oligopoly competition games is to enable costly targeted monetary punishment. There is a broad literature on public good games that shows that the ability to punish leads to more cooperation (see, e.g., Ostrom et al. (1992), Fehr and Gächter (2000), Fehr and Gächter (2002), Güerker et al.

---

<sup>1</sup>See, e.g., Trkman and Desouza (2012) for a more detailed analysis of the potential risks.

(2006), Gächter et al. (2008), Nikiforakis and Normann (2008), Reuben and Riedl (2013), Gächter et al. (2017)). Further, there is similar evidence for prisoner's dilemma games (see, e.g., Dreber et al. (2008) and Bone et al. (2015)) and oligopoly competition games (see Roux and Thöni (2015)), indicating that more control, i.e., the ability to punish, generally increases cooperation. This connection breaks down only in a few exceptional cases. For example, Nikiforakis and Normann (2008) show that the effectiveness of the punishment tool is important, and Nikiforakis (2008) shows that the appearance of counter-punishments can cause a negative effect of punishment on cooperation. However, it is also important to point out that the overall effect of monetary punishment is ambivalent. Many studies show that the costs caused by punishment can outweigh the positive effect of enhanced cooperation, resulting in a negative net effect on overall profits (see, e.g., Ostrom et al. (1992), Fehr and Gächter (2002), Dreber et al. (2008), Gächter et al. (2008), Gächter et al. (2017)).

In our context, firms cannot directly punish other firms for not sharing data. Rather, depending on the level of control, they can only revoke data access to some or all firms. This is more closely related to previous studies that consider non-monetary punishments in the form of ostracism. Especially with regard to public good games, many studies have found that the possibility to ostracize non-cooperators is sufficient to increase contributions to the public good (see, e.g., Cinyabuguma et al. (2005), Güth et al. (2007) or Maier-Rigaud et al. (2010)). This is consistent with the result of Riedl and Ule (2002) in the context of an n-person prisoner's dilemma experiment. In other words, to encourage cooperation, it seems sufficient to provide control over who to cooperate with.

This is confirmed by the results of experiments comparing bilateral relationships with multilateral relationships. To see the connection, consider that bilateral relationships are characterized by a firm being able to selectively decide whether or not to cooperate with a particular other firm. In a multilateral relationship, a firm cannot discriminate, i.e., it cooperates with all other firms or with none. For n-entity prisoner's dilemma games, Marwell and Schmitt (1972) find in their experiment that the cooperation is significantly higher in a

game with a bilateral relationship than in a game with a trilateral relationship. In a more recent experiment, Nosenzo et al. (2015) find the same result for a public good game when comparing a game with a bilateral relationship with a game with three, four or eight entities. Further, this is in line with evidence from oligopoly experiments with Cournot competition. Dolbear et al. (1968), Huck et al. (2004) and Orzen (2008) find markets with two firms to be more collusive than markets with more than two firms. This result is confirmed for Bertrand and Cournot competition by Horstmann et al. (2018) by means of a meta-analysis and additional experiments.

Moreover, the literature provides evidence that it is not only important whether a firm can control with whom it cooperates, but also on the degree of the cooperation. Such a more differentiated type of control allows for more differentiated non-monetary punishment, which can enhance cooperation, because the punishment can take the reasons and circumstances for deviation into account (Ostrom, 1990). In this vein, Wright (2013) investigates which punishment strategies are successful under a Bertrand competition. He finds that tacit collusion between competitors works best in markets where firms punish a small price cut of a competitor proportionally with a similarly small price cut in the next round, rather than with a disproportionately large price cut. Accordingly, a disproportional strategy was used by only a few participants, while the majority used a proportional strategy (Wright, 2013).

In summary, previous literature suggests that cooperation can be enhanced by giving firms more control over who they cooperate with and cooperation can be further enhanced by allowing for fine-granular responses based on the severity of a cooperation partner's deviation.

**Transparency:** In terms of transparency, a distinction is often made between information about the actions and information about individual profits of other firms (Fiala and Suetens, 2017; Nikiforakis, 2010; Altavilla et al., 2006). The public goods literature consistently shows that full information about others' *contributions* leads to more cooperation than less

information (Fiala and Suetens, 2017; Nikiforakis, 2010; Sell and Wilson, 1991). This means that group members who receive information about individual contributions in a public good game tend to invest more than those who receive no or only aggregated information. On the other hand, full information about others' *profits* leads to less cooperation (Fiala and Suetens, 2017; Nikiforakis, 2010). More so, Nikiforakis (2010) finds less cooperation in a public good experiment (in which subjects can impose a costly targeted punishment) when payoffs and contributions are known compared to a situation in which only contributions are known; but more cooperation compared to a situation in which only profits are known.

In our experiment, information on data contributions and information on profits are inevitably linked, because all firms are ex-ante symmetric and thus, each firm can infer profits from data contributions. Thus, no clear prediction can be derived from the above literature on the effect of transparency. In this regard, our setting is more similar to previous experiments on oligopoly competition, which has considered the joint effect of information about competitors' actions and profits on collusion (Huck et al., 1999, 2000; Altavilla et al., 2006). However, this literature also does not find a clear effect. While it was found that aggregated information on individual profits and quantities leads to more collusion than detailed information in Cournot markets (Huck et al., 1999, 2000), no such difference was found in Bertrand markets (Altavilla et al., 2006).

Also the literature on coordinated cooperation and prisoner's dilemma games highlights that the influence of information on cooperation strongly depends on the respective setting of the game. In this context, it has been studied how the probability with which the action of the other player is made visible before making a decision facilitates cooperation. Unakafov et al. (2020) and Unakafov et al. (2019) show that action visibility stimulates cooperation in coordinated cooperation games, but leads to less cooperation in iterated prisoner's dilemma games.

Taken together, previous experimental studies did not find a clear impact of transparency on cooperation. Rather, the literature suggests that the direction of impact depends on



the type of information and the design of institutions (Ostrom, 2000). This also means that we cannot derive a meaningful prediction on the relationship between information and cooperation identified in this literature for the new context of data sharing studied here.

## 3 Experimental Design

### 3.1 The Model

We develop a simple game-theoretical model in order to capture the idiosyncratic strategic trade-offs that are inherent to B2B data sharing. This will then provide the theoretical framework for our experimental analysis. We consider a set of  $n$  firms, which can, in principle, share data with each other. Each firm is endowed with a set of (unique) data points. The firms can simultaneously decide how much of their data to share. We then make the following key assumptions on the strategic interactions in the context of B2B data sharing:

- (A1) The total profit of all firms is the higher, the more data the firms share.
- (A2) The profit of a firm is the higher, the more data the other firms share with it.
- (A3) The competitive advantage of a firm is the greater, the more data the other firms share with it.

These assumptions reflect that each firm would profit from having access to more data (e.g., because it allows for data-driven innovations), but that sharing of data also bears indirect costs through the (potential) competitive advantage that the data conveys to the rival firms (i.e. the sharing partners). However, reflecting the believe of many policymakers, such as the European Commission, the economy as a whole (in terms of producers' surplus) would be better off if more data is shared between the firms (European Commission, 2020). All these

assumptions are captured by the following stylized profit function of a firm  $i \in \{1, \dots, n\}$ :

$$\Pi_i = m x_i + m_s \sum_{j=1; i \neq j}^n s_{ji} x_j - \frac{c}{n-1} \left[ \sum_{j=1; i \neq j}^n \left( m x_j + m_s \sum_{k=1; k \neq j}^n s_{kj} x_k \right) \right] \quad (1)$$

where  $x_i$  is firm  $i$ 's data endowment,  $s_{ij} \in [0, 1]$  is the percentage of firm  $i$ 's data endowment that it shares with firm  $j$ ,  $m$  is the revenue per unit from own data,  $m_s$  is the revenue per unit from data shared by other firms and  $c$  is the degree of competition with  $c \in [0, 1]$ . Thus, the first term in the profit function represents the revenue from the firm's own data; the second term represents the revenue from data shared by the other firms with firm  $i$ ; and the third term covers the competitive influence of other firms depending on the average amount of data to which they have access. Specifically, we model the profit of a firm as the revenue from the data available to it minus  $c$  times the average revenue of its competitors from the data available to them. For the experiment we used the following parameter specifications:  $n = 3$ ,  $m = m_s = 1$ ,  $x_i = 100 \forall i$  and  $c = \frac{1}{3}$ , which relate to a market with three firms and a medium level of competition.

In the experiment, the game will be repeated  $n$  times with a fixed ending. Note that in this case, for any  $c > 0$ , the unique subgame perfect Nash equilibrium is for each firm not to share any data in any period. In this sense, our model bears a similar theoretical benchmark as in i) standard public good games, where each player should not contribute to the public good in equilibrium, ii) prisoner's dilemma games, where each player should defect in equilibrium, and iii) oligopoly competition games, where each firm should play the one-shot static Nash equilibrium in each round, instead of colluding. However, for all these games it is also well known that players nevertheless show a certain degree of cooperation in experiments. Consequently, we also expect to observe a degree of cooperation in the case of data sharing that exceeds the theoretical prediction that no data is shared. However, as argued above, we also expect that degree of data sharing will depend on the degree of control and transparency that differ across treatments, which we explain next.

## 3.2 Design and Treatments

The experiment is based on a 3x2 full factorial design, in which we systematically consider three levels of control and two levels of transparency in six different treatments.

At the lowest level of control, called *Collective Sharing*, a firm can determine how much of its data endowment it wants to share, but has no differentiated control over who accesses its shared data. That is, each firm can only share the same percentage of its data with all other firms. This reflects situations where a firm uploads part of its data to a data pool that provides unrestricted data access to all participating firms. At a medium level of control, called *Excludable Sharing*, each firm can take the same actions as under Collective Sharing, but may additionally exclude some firms from accessing its shared data. That is, a firm cannot differentiate the percentage of data shared with each other firm, but can deny some firms access to its data altogether. This reflects situations where a data pool allows firms to decide who can access their uploaded data, or where the firm licenses interested parties before granting access to its public API. At the highest level of control, called *Individual Sharing*, a firm can take the same actions as under Excludable Sharing, but is additionally able to decide independently for each data recipient exactly how much of its data can be accessed, i.e., firms can set the amount of data to be shared specifically for each firm. This reflects a situation where a firm can engage in individual sharing contracts, or can grant fine-granular access rights for private APIs.

With respect to transparency, we distinguish between full information and partial information settings. Under *Full Information*, each firm can observe every previous data sharing transaction by all other firms, even if it relates to transactions in which it was not involved (as the sharing or the receiving firm). Thus, each firm knows exactly which firm has shared how much data with each other firm and how much profit it has made. By contrast, under *Partial Information* each firm only has knowledge on previous transactions in which it was involved, but not on transactions that involve the two other firms only. That is, in the case of Excludable Sharing and Individual Sharing firms only know how much data each firm has

shared with them (and, of course, how much data they shared with each other firm), but not how much data the other firms have shared with each other. Under Collective Sharing all firms know that each firm must share the same percentage with each other firm. Thus, here *Partial Information* means that each firm only has aggregated information on the other firms total contributions, but not their individual contributions. That is, each firm can observe how much data has been contributed to the data pool, but they do not know which firm has contributed how much of it.

An overview of all six treatments and their abbreviations used in the following is provided by Table 1. Taken together, these allow us to study the various degrees of control and transparency, but also the possible interactions between them, with respect to their ability to elicit more data sharing.

<b>Control dimension</b>	<b>Transparency dimension</b>	
	Partial Information (P)	Full Information (F)
Collective Sharing	<i>Collective<sub>P</sub></i>	<i>Collective<sub>F</sub></i>
Excludable Sharing	<i>Excludable<sub>P</sub></i>	<i>Excludable<sub>F</sub></i>
Individual Sharing	<i>Individual<sub>P</sub></i>	<i>Individual<sub>F</sub></i>

Table 1: Overview of the experimental treatments.

## 4 Procedures

In total there were 24 sessions, 4 sessions for each treatment, with a total of 354 participants. All sessions were conducted at the School of Business, Economics and Information Systems at the University of Passau, Germany from January to July 2020. Subjects were recruited from the student subject pool of the university using ORSEE (Greiner, 2015). Students are considered to be well-suited surrogates for managerial decision making with respect to internal validity of the experiments (Graf-Vlachy, 2019), and it is therefore common to employ student subjects also in firm experiments (see, e.g., Brandts and Potters, 2018; Holt, 2020). The experiment was computerized with *oTree*, a software platform for economic experiments

(Chen et al., 2016). In a session 15 participants were matched randomly and anonymously to groups of three firms each, with each participant representing one firm.<sup>2</sup> Each group represents an independent market, and thus each group played the game (see Section 3.1) over a total of 30 rounds with the same participants in each round (fixed partner matching). In all treatments, subjects were fully informed about the timeline of the experiment and the consequences of their actions.

A session comprised seven steps. First, participants were randomly assigned to a seat. Subjects could neither see nor communicate with each other. Throughout the experiment, participants could take notes on a notepad provided for them. Second, the experimental instructions were handed out in print and read aloud from a recording. The recording ensures that the voice, accent and emphasis were identical in each session. Third, before starting the actual experiment, the participants had to answer a series of comprehension questions. They could only proceed if all questions were answered correctly. Fourth, participants were shown a video explaining the user interface of the experimental software. The video clearly illustrated the processes in the experiment and familiarized participants with the user interface through which they could submit their decision on how much data to share with the other firms, and in which they were presented information on past actions, as well as a profit calculator. Fifth, participants proceeded to the first round of the game without prior practice rounds in order to prevent unobserved learning effects. Throughout the game, participants were able to access the profit calculator which allowed them to check the consequences for data availability and profits for each firm, for every possible constellation of the decisions taken by the three firms. Once the participants had made their decision, they were redirected to a waiting page. On the waiting page participants were still able to use the profit calculator. After all participants have made their decision for this round, participants were presented with the results of the current round. That is, they were shown the decisions, profits and available data from the current round with the level of detail that corresponds to their respective treatment. Across

---

<sup>2</sup>In two sessions, the number of participants was reduced to 12 due to no shows.

all treatments, it took about an hour to complete the 30 rounds of playing the game. Sixth, at the end of the game, the subjects had to answer some post-questions. Finally, the subjects were paid out the profit of his or her firm from exactly one of the 30 rounds. The payout of the profit from exactly one round avoids budget effects. Each participant determined the round to be paid out by throwing a 30-sided dice. The profit their firm had made at the end of this round (measured in Thalers) was converted into Euro; 10 Thalers correspond to one Euro. Finally, the profit from the selected round in addition to a participation fee of 10 Euro was paid to the subjects privately and in cash. The duration of a session from entering the laboratory to the end of payment was about two hours. The average payout was 24.06 Euro.

## 5 Data Analysis and Results

Our analysis focuses on how much data is shared in a given market. In the following we consider the average amount of data shared (which we also refer to as ‘data contributions’) in a given group in a single round as the lowest unit of observation. Note that at this level, each observation already aggregates up to six individual decisions (six under Excludable Sharing and Individual Sharing, but three under Collective Sharing), as each of the three firms in a given market has to choose a data contribution for each of the other two firms (but each firm chooses only one data contribution under Collective Sharing).<sup>3</sup> We conduct parametric tests comparing differences between treatments, done at the level of statistically independent observations at the market (group) level (118 observations in total), which we derive by averaging observations relating to a given market (group) over rounds.<sup>4</sup> In order

---

<sup>3</sup>Note that, as there are always three firms in a given market, the average amount of data shared is perfectly correlated with the total amount of data shared in a given market, and thus, for the statistical analysis it is immaterial whether we consider the average or the total amount of data shared in a given market. We chose to consider the average contribution for our analysis, because it has the same range as the individual contributions (between 0 and 100), which eases comprehension.

<sup>4</sup>Our results are robust to the use of non-parametric tests, but since these tests generally lack power over parametric counterparts and cannot deal with the data as efficiently, we choose to report only parametric test results.

to avoid distortions in the averages from learning effects in the first few rounds and from endgame effects in the last few rounds, we do not take the first two and the last two rounds into account when computing the averages over rounds.<sup>5</sup> Table 2 provides the summary statistics of the statistically independent observations at the market level and Figure 1 provides a histogram of these observations, i.e., of the average data contribution in a given market over all rounds (3-28), ranging from 0 (no firm in the market has shared any data in any round) to 100 (all firms in the market have shared all their data in all rounds).

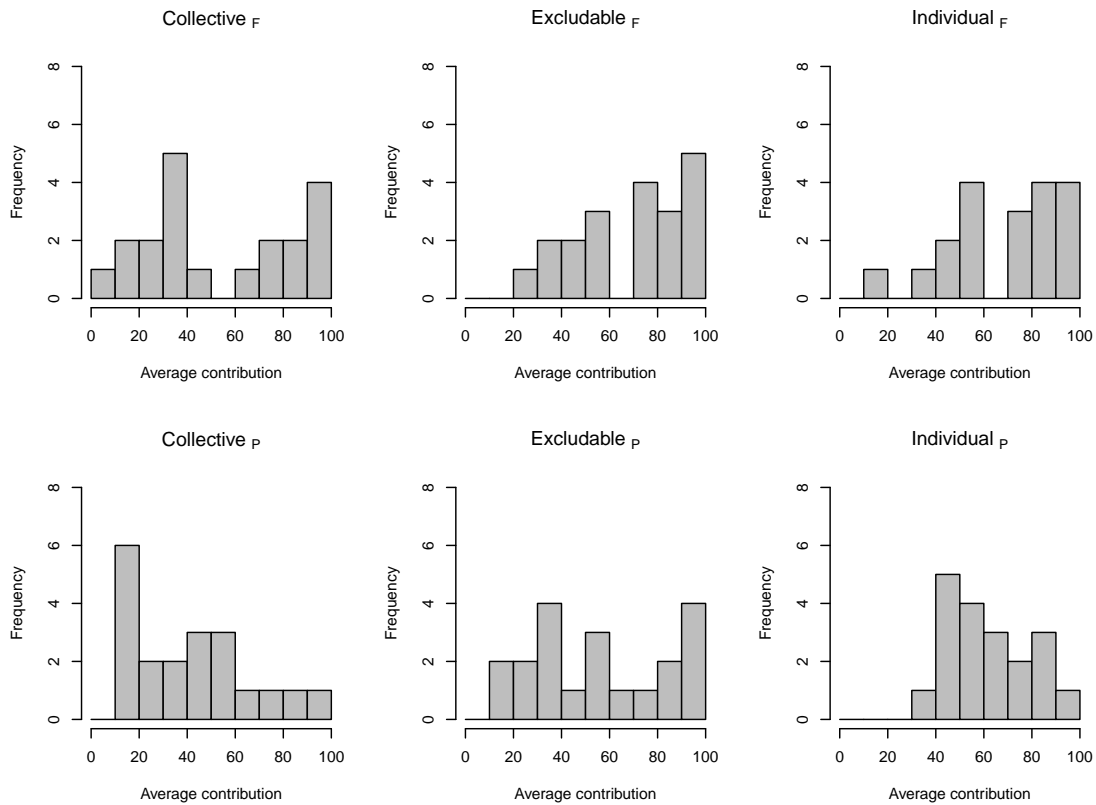


Figure 1: Histogram of the average data shared in each market over all rounds for each treatment.

**Analysis of main effects:** In order to test for main effects of control and transparency on the average data contribution in a market, we perform a one-way analysis of variance (ANOVA) to test for equality of means in the levels of the control and transparency dimen-

<sup>5</sup>Our results do not depend on the exclusion of the first and last rounds.

Treatment	Obs.	Mean	SD
<i>Collective<sub>F</sub></i>	20	54.53	33.17
<i>Excludable<sub>F</sub></i>	20	69.53	23.49
<i>Individual<sub>F</sub></i>	19	67.08	23.68
<i>Collective<sub>P</sub></i>	20	41.31	24.77
<i>Excludable<sub>P</sub></i>	20	56.66	29.17
<i>Individual<sub>P</sub></i>	19	62.62	17.59

Table 2: Summary statistics of the average data shared in each market for each treatment averaged over all rounds (3-28).

sion. The results indicate that there is a significant difference in the main effect control ( $F(2, 115) = 5.025, p = .01, \eta_p^2 = .08$ ) and a significant difference in the main effect transparency ( $F(1, 116) = 4.398, p = .04, \eta_p^2 = .04$ ).

To check which levels of control differ significantly, we perform a post-hoc Tukey-Test.<sup>6</sup> The results show, that in the control dimension Collective Sharing differs significantly from Excludable Sharing ( $p = .03$ ) and from Individual Sharing ( $p = .01$ ). Under Individual Sharing an average of 16.9 data points more were shared than under Collective Sharing and under Excludable Sharing an average of 15.2 data points more were shared than under Collective Sharing. However, we do not find a significant difference between Excludable Sharing and Individual Sharing ( $p = .95$ ).

**Insight 1.** *Medium control over data access (Excludable Sharing) leads to the same level of data sharing as high control over data access (Individual Sharing). However, low control over data access (Collective Sharing) yields significantly less data sharing than medium or high control.*

**Insight 2.** *High transparency about data contributions (Full Information) leads to significantly more data sharing than low transparency (Partial Information).*

**Analysis of the interaction effect:** Next, we perform a two-way analysis of variance to test for a significant interaction effect between control and transparency. Control includes

<sup>6</sup>The transparency dimension consists of only two characteristics and therefore no post-hoc test is necessary.



three levels (Collective, Excludable and Individual) and transparency includes two levels (Full and Partial). The null hypothesis is that there is no interaction between the control and transparency dimension, i.e.  $H_0: Control \times Transparency \text{ interaction} = 0$  (Kirk, 2009, p. 39). The results show that the null hypothesis cannot be rejected ( $F(2, 112) = .36, p = .70, \eta_p^2 = .01$ ).

In our previous analysis we have averaged observations over all rounds and thus ignored how data sharing in markets may evolve over time. Indeed, the average data contributions in a given market increase over the rounds ( $r_s = .12, p < .01$ ) and Table 3 suggests that the firms' propensity to share data in early rounds (rounds 3-15), when market relationships are still relatively short-lived, seems to differ from that in late rounds (rounds 16-28), when the market relationships are more long-lived.

Treatment	Early Rounds (3-15)			Late Rounds (16-28)		
	Obs.	Mean	SD	Obs.	Mean	SD
<i>Collective<sub>F</sub></i>	20	53.2	30.3	20	55.9	37.9
<i>Excludable<sub>F</sub></i>	20	62.3	25.0	20	76.8	24.0
<i>Individual<sub>F</sub></i>	19	60.6	21.8	19	73.6	27.4
<i>Collective<sub>P</sub></i>	20	40.0	23.0	20	42.6	29.1
<i>Excludable<sub>P</sub></i>	20	53.0	28.8	20	60.3	30.9
<i>Individual<sub>P</sub></i>	19	53.8	18.6	19	71.5	19.1

Table 3: Summary statistics for independent observations in early and late rounds.

To compare the average data contributions in early and late rounds, we perform a dependent-samples t-test across all treatment conditions. The results show that there is a significant difference in the average data contributions for early ( $M = 53.8, SD = 25.5$ ) and late rounds ( $M = 63.3, SD = 30.6$ );  $t(117) = -6.51, p < .001$ , indicating that the duration of the market relationship does indeed have an effect on data contributions.

In order to test in more detail for which levels of control and transparency early and late rounds differ significantly, we perform a t-test for each treatment separately. The results show that for both transparency levels, data contributions differ significantly between early and late rounds in the control dimensions Excludable Sharing ( $Excludable_F: t(19) = -4.59, p$

$< 0.001$ , Excludable<sub>P</sub>:  $t(19) = -2.57$ ,  $p = 0.02$ ) and Individual Sharing (Individual<sub>F</sub>:  $t(18) = -3.89$ ,  $p = 0.001$ , Individual<sub>P</sub>:  $t(18) = -5.70$ ,  $p < 0.001$ ), but not in Collective Sharing (Collective<sub>F</sub>:  $t(19) = -0.68$ ,  $p = 0.51$ , Collective<sub>P</sub>:  $t(19) = -0.66$ ,  $p = 0.52$ ).

Specifically, our results suggest that under Excludable and Individual Sharing, for both transparency levels, on average more data is shared in long-lived market relationships than in short-lived relationships.

**Insight 3.** *With medium and high control over data access (Excludable and Individual Sharing), in long-lived market relationships (late rounds) on average more data is shared than in short-lived market relationships (early rounds).*

## 6 Discussion

A central insight of our experimental results is that, albeit more control generally yields a (weakly) higher level of data sharing, a medium level of control (Excludable Sharing) may already be sufficient to increase a firm's propensity to share data on B2B data sharing platforms (Insight 1). Excludable Sharing corresponds to a technical and contractual design where firms provide a common, standardized interface to their data stream (e.g. through APIs defined by the B2B sharing platform), but retain control over the scope and scale of data they share through this interface, and the right to revoke and grant access privileges individually for any other firm. However, they do not have the possibility to discriminate the scope or scale of access for individual firms beyond that. This allows for very efficient and lean technical and legal implementations of B2B data sharing platforms. By contrast, Individual Sharing, where the scope and scale of each sharing relationship is negotiated individually, is a much less efficient design whose complexity potentially grows quadratically with the number of sharing partners; but, as our results indicate, albeit being more complex and costly to implement and operate, Individual Sharing does not necessarily provide additional benefits for the viability and prosperity of a B2B data sharing platform. In reverse, our findings also

provide a clear insight that Collective Sharing, that is feeding firms' data into a common data pool without the possibility to control which other firms have access to it, reduces each firm's propensity to share data in the first place. This is in line with previous insights from the public goods literature, which finds that the possibility of exclusion of non-cooperators increases contributions (see, e.g. Cinyabuguma et al., 2005; Güth et al., 2007; Maier-Rigaud et al., 2010). Accordingly, a design where all firms have unrestricted access to shared data seems to be dominated by designs where firms have at least control over who exactly can access their shared data.

Further, transparency about the data sharing behavior of *others* is an important antecedent for encouraging more data contributions (Insight 2). This echoes some previous experimental findings that were derived in the contexts of public good games and oligopolies (see, e.g. Fiala and Suetens, 2017; Sell and Wilson, 1991), albeit the literature also suggests that the effect of transparency depends on environmental factors (Ostrom, 2000; Møllgaard and Overgaard, 2001), and thus should not be carelessly generalized to different contexts. In our context, it could be argued that transparency on all data transactions on a B2B sharing platform fosters confidence in the system as a whole, as well as in other sharing partners, as it allows to observe *all* transactions on the platform and in particular a sharing partner's behavior also in other relationships.

However, confidence in the system and the partners also naturally evolves over time. This is consistent with our finding that with medium and high level of data access control (Excludable and Individual Sharing), for both levels of transparency (low and high transparency), more data is shared in long-lived market relationships than in short-lived relationships (Insight 3). This is an important insight especially when setting up B2B sharing platforms, where firms that do not necessarily operate in the same (product) markets and have therefore not established a prior working relationship are now supposed to share data with each other.

Finally, our results also suggest that B2B sharing platforms can partially substitute control with transparency, or the other way around, as we find that both, transparency and control are - independently of each other - conducive to more data sharing. This possibility to balance between transparency and control can be very important in practice, depending on the nature of the data and technical considerations. For example, legal restrictions may limit the extent of transparency, i.e., limit the information that can be shared about data trades that relate to others. In this case, a finer control architecture (Excludable Sharing or Individual Sharing) should be chosen to increase the firm's propensity to share data. In other cases, it might be that finer access rights and the possibility for individual contracting on data sharing, may considered to be too burdensome from a technical or legal perspective. In this case, a high level of transparency should be chosen to stimulate data contributions nevertheless.

## **7 Conclusions and Directions for Future Research**

In this paper, we have identified firms' data access control and transparency about previous data transactions as important dimensions that should be considered when designing B2B data sharing platforms. Based on an economic laboratory experiment and an abstract data sharing scenario, we systematically vary the levels of control and transparency in a B2B data sharing platform, and identify nuanced effects that should be considered by those B2B sharing platforms that are currently struggling to attract contributing firms. In particular, our results suggest that a medium level of control, which allows a firm to share data with all other firms, but also allow the firms to exclude individual firms from access, paired with full transparency about all data transactions on the platform, is an efficient design that is at the same time conducive to a high level of data sharing.

In closing, we wish to point to a number of limitations of our study that may prompt future research. The strength of the experimental methodology lies in its internal validity.

That is, the experimental approach allowed us to systematically vary the control and transparency dimensions, while keeping environmental conditions constant. This would not have been possible in the field, where environmental conditions change constantly. However, on the flip side, high internal validity comes at the cost of external validity. One cannot be sure that the results obtained in the experiment carry over to the field. Yet, laboratory experiments can provide actionable insights for practice (for a discussion see Plott, 1987), and fulfill a similar purpose as a wind tunnel in engineering: If a design already fails in the lab, why should we expect it to perform well in the field? Nevertheless, if feasible, experimentation of different designs done directly at a B2B data sharing platform would be an important next step for future research.

The real world is, of course, more complex than the abstract decision situation that we have presented to subjects in our experiment. While such abstraction is necessary, other routes could have been taken in doing so. For example, we chose not to differentiate between the amount and the value of data, and we have not placed the data in a particular context in order to avoid framing effects. However, future research could consider alternative abstractions of B2B data sharing scenarios, and study other dimensions of B2B data sharing contexts. For example, how the framing of the purpose of data sharing affects firms' propensity to share, or which strategy firms pursue if they can share low value or high value data. Similarly, in our setting the degree of competition between firms was kept constant ( $c = 1/3$ ), but future research may also consider how the degree of competition between firms affects data contributions, or whether it interacts with the transparency or control dimension.

## References

- Altavilla, C., Luini, L., and Sbriglia, P. (2006). Social learning in market games. *Journal of Economic Behavior & Organization*, 61(4):632–652.
- BMW, F. (2020a). Gaia-x: A federated data infrastructure for europe. <https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html>. Accessed November 18, 2021.
- BMW, F. (2020b). Gaia-x: Technical architecture. Available at <https://www.bmw.de/Redaktion/EN/Publikationen/gaia-x-technical-architecture.pdf>.
- Bogers, M. (2011). The open innovation paradox: knowledge sharing and protection in r&d collaborations. *European Journal of Innovation Management*, 14(1):93–117.
- Bone, J. E., Wallace, B., Bshary, R., and Raihani, N. J. (2015). The effect of power asymmetries on cooperation and punishment in a prisoner’s dilemma game. *PloS One*, 10(1):e0117183, 1–15.
- Brandts, J. and Potters, J. (2018). Experimental industrial organization. In *Handbook of Game Theory and Industrial Organization, Volume II*. Edward Elgar Publishing.
- Cabrera, A. and Cabrera, E. F. (2002). Knowledge-sharing dilemmas. *Organization Studies*, 23(5):687–710.
- Chen, D. L., Schonger, M., and Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.
- Cinyabuguma, M., Page, T., and Putterman, L. (2005). Cooperation under the threat of expulsion in a public goods experiment. *Journal of Public Economics*, 89(8):1421–1435.

- De Prieëlle, F., De Reuver, M., and Rezaei, J. (2020). The role of ecosystem data governance in adoption of data platforms by internet-of-things data providers: Case of dutch horticulture industry. *IEEE Transactions on Engineering Management*, pages 1–11.
- Dolbear, F. T., Lave, L. B., Bowman, G., Lieberman, A., Prescott, E., Rueter, F., and Sherman, R. (1968). Collusion in oligopoly: an experiment on the effect of numbers and information. *The Quarterly Journal of Economics*, 82(2):240–259.
- Dreber, A., Rand, D. G., Fudenberg, D., and Nowak, M. A. (2008). Winners don't punish. *Nature*, 452(7185):348–351.
- European Commission (2020). Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions "a european strategy for data". Communication COM(2020) 66 final, European Commission, Brussels.
- Fehr, E. and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994.
- Fehr, E. and Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868):137–140.
- Fiala, L. and Suetens, S. (2017). Transparency and cooperation in repeated dilemma games: a meta study. *Experimental Economics*, 20(4):755–771.
- Gächter, S., Mengel, F., Tsakas, E., and Vostroknutov, A. (2017). Growth and inequality in public good provision. *Journal of Public Economics*, 150:1–13.
- Gächter, S., Renner, E., and Sefton, M. (2008). The long-run benefits of punishment. *Science*, 322(5907):1510–1510.

- Ghosh, A., Mahdian, M., McAfee, R. P., and Vassilvitskii, S. (2015). To match or not to match: Economics of cookie matching in online advertising. *ACM Transactions on Economics and Computation (TEAC)*, 3(2):1–18.
- Graf-Vlachy, L. (2019). Like student like manager? using student subjects in managerial debiasing research. *Review of Managerial Science*, 13(2):347–376.
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125.
- Gürerk, Ö., Irlenbusch, B., and Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science*, 312(5770):108–111.
- Güth, W., Levati, M. V., Sutter, M., and Van Der Heijden, E. (2007). Leading by example with and without exclusion power in voluntary contribution experiments. *Journal of Public Economics*, 91(5-6):1023–1042.
- Hall, D. W. (2020). Growing the artificial intelligence industry in the uk. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/652097/Growing\\_the\\_artificial\\_intelligence\\_industry\\_in\\_the\\_UK.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf). Accessed July 19, 2021.
- Holt, C. A. (2020). 5. industrial organization: A survey of laboratory research. In *The handbook of experimental economics*, pages 349–444. Princeton University Press.
- Horstmann, N., Krämer, J., and Schnurr, D. (2018). Number effects and tacit collusion in experimental oligopolies. *The Journal of Industrial Economics*, 66(3):650–700.
- Huang, M.-H., Wang, E. T., and Seidmann, A. (2007). Price mechanism for knowledge transfer: An integrative theory. *Journal of Management Information Systems*, 24(3):79–108.



- Huck, S., Normann, H.-T., and Oechssler, J. (1999). Learning in cournot oligopoly—an experiment. *The Economic Journal*, 109(454):80–95.
- Huck, S., Normann, H.-T., and Oechssler, J. (2000). Does information about competitors' actions increase or decrease competition in experimental oligopoly markets? *International Journal of Industrial Organization*, 18(1):39–57.
- Huck, S., Normann, H.-T., and Oechssler, J. (2004). Two are few and four are many: number effects in experimental oligopolies. *Journal of Economic Behavior & Organization*, 53(4):435–446.
- Jarke, M., Otto, B., and Ram, S. (2019). Data sovereignty and data space ecosystems. *Business & Information Systems Engineering*, 61:549–550.
- Jentzsch, N., Sapi, G., and Suleymanova, I. (2013). Targeted pricing and customer data sharing among rivals. *International Journal of Industrial Organization*, 31(2):131–144.
- Kembro, J., Selviaridis, K., and Näslund, D. (2014). Theoretical perspectives on information sharing in supply chains: a systematic literature review and conceptual framework. *Supply Chain Management*, 19(5/6):609–625.
- Kim, K. K., Umanath, N. S., and Kim, B. H. (2005). An assessment of electronic information transfer in b2b supply-channel relationships. *Journal of Management Information Systems*, 22(3):294–320.
- Kirk, R. E. (2009). *The SAGE Handbook of Quantitative Methods in Psychology*. SAGE, Los Angeles, London, New Delhi, Singapore, Wahsington DC.
- Krämer, J., Schnurr, D., and Wohlfarth, M. (2019). Winners, losers, and facebook: The role of social logins in the online advertising ecosystem. *Management Science*, 65(4):1678–1699.
- Lee, H. L. and Whang, S. (2000). Information sharing in a supply chain. *International Journal of Manufacturing Technology and Management*, 1(1):79–93.

- Loebbecke, C., van Fenema, P. C., and Powell, P. (2016). Managing inter-organizational knowledge sharing. *The Journal of Strategic Information Systems*, 25(1):4–14.
- Maier-Rigaud, F. P., Martinsson, P., and Staffiero, G. (2010). Ostracism and the provision of a public good: experimental evidence. *Journal of Economic Behavior & Organization*, 73(3):387–395.
- Martens, B., de Streel, A., Graef, I., Tombal, T., and Duch-Brown, N. (2020). Business-to-business data sharing: An economic and legal analysis. <https://www.econstor.eu/bitstream/10419/227701/1/1726726282.pdf>. EU Science Hub. Working Paper.
- Marwell, G. and Schmitt, D. R. (1972). Cooperation in a three-person prisoner’s dilemma. *Journal of Personality and Social Psychology*, 21(3):376.
- Microsoft (2019). The open data initiative. <https://www.microsoft.com/en-us/open-data-initiative>. Accessed May 26, 2021.
- Ministry of Economic Affairs and Climate Policy (2019). Dutch digitalisation strategy: Dutch vision on data sharing between businesses. <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&ved=2ahUKEwj2kd0qmc3jAhVFD0wKHdieBb4QFjADegQIBBAC&url=https%3A%2F%2Fwww.government.nl%2Fbinaries%2Fgovernment%2Fdocuments%2Freports%2F2019%2F02%2F01%2Fdutch-vision-on-data-sharing-between-businesses%2FDutch%2Bvision%2Bon%2Bdata%2Bsharing%2Bbetween%2Bbusinesses.pdf&usg=AOvVaw10VaXErN1EgRhDZD8h19Gi>. Accessed May 26, 2021.
- Møllgaard, H. P. and Overgaard, P. B. (2001). Market transparency and competition policy. <https://www.econstor.eu/bitstream/10419/208446/1/cbs-wp2001-06.pdf>. Working Paper.

- Natwichai, J., Li, X., and Orłowska, M. (2005). Hiding classification rules for data sharing with privacy preservation. In *International Conference on Data Warehousing and Knowledge Discovery*, pages 468–477. Springer.
- Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, 92(1-2):91–112.
- Nikiforakis, N. (2010). Feedback, punishment and cooperation in public good experiments. *Games and Economic Behavior*, 68(2):689–702.
- Nikiforakis, N. and Normann, H.-T. (2008). A comparative statics analysis of punishment in public-good experiments. *Experimental Economics*, 11(4):358–369.
- Norman, P. M. (2001). Are your secrets safe? knowledge protection in strategic alliances. *Business Horizons*, 44(6):51–61.
- Nosenzo, D., Quercia, S., and Sefton, M. (2015). Cooperation in small groups: the effect of group size. *Experimental Economics*, 18(1):4–14.
- Orzen, H. (2008). Counterintuitive number effects in experimental oligopolies. *Experimental Economics*, 11(4):390–401.
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.
- Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives*, 14(3):137–158.
- Ostrom, E., Walker, J., and Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *American Political Science Review*, 86(2):404–417.
- Otto, B. (2019). Interview with reinhold achatz on “data sovereignty and data ecosystems”. *Business & Information Systems Engineering*, 61(5):635–636.

- Pauer, A., Nagel, L., Fedkenhauser, T., Yvonne, F.-S., and Resetko, A. (2018). Data exchange as a first step towards economy. <https://www.pwc.de/en/digitale-transformation/data-exchange-as-a-first-step-towards-data-economy.pdf>. Accessed July 12, 2021.
- Plott, C. R. (1987). Dimensions of parallelism: Some policy applications of experimental methods. *Laboratory experimentation in economics: Six points of view*, pages 193–219.
- Reuben, E. and Riedl, A. (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, 77(1):122–137.
- Riedl, A. and Ule, A. (2002). Exclusion and cooperation in social network experiments. [https://www.researchgate.net/profile/Arno\\_Riedl/publication/237666617\\_Exclusion\\_and\\_Cooperation\\_in\\_Social\\_Network\\_Experiments/links/02e7e52d48a2e40046000000/Exclusion-and-Cooperation-in-Social-Network-Experiments.pdf](https://www.researchgate.net/profile/Arno_Riedl/publication/237666617_Exclusion_and_Cooperation_in_Social_Network_Experiments/links/02e7e52d48a2e40046000000/Exclusion-and-Cooperation-in-Social-Network-Experiments.pdf). CREED, Faculty of Economics and Econometrics, University of Amsterdam. Working Paper.
- Roux, C. and Thöni, C. (2015). Collusion among many firms: The disciplinary power of targeted punishment. *Journal of Economic Behavior & Organization*, 116:83–93.
- Sell, J. and Wilson, R. K. (1991). Levels of information and contributions to public goods. *Social Forces*, 70(1):107–124.
- Telekom (2019). Data intelligence hub. <https://dih.telekom.net/en/>. Accessed May 26, 2021.
- Trkman, P. and Desouza, K. C. (2012). Knowledge risks in organizational networks: An exploratory framework. *The Journal of Strategic Information Systems*, 21(1):1–17.

- Unakafov, A. M., Schultze, T., Gail, A., Moeller, S., Kagan, I., Eule, S., and Wolf, F. (2020). Emergence and suppression of cooperation by action visibility in transparent games. *PLoS Computational Biology*, 16(1):e1007588, 1–32.
- Unakafov, A. M., Schultze, T., Kagan, I., Moeller, S., Gail, A., Treue, S., Eule, S., and Wolf, F. (2019). Evolutionary successful strategies in a transparent iterated prisoner’s dilemma. In *International Conference on the Applications of Evolutionary Computation (Part of EvoStar)*, pages 204–219. Springer.
- Wang, S. and Noe, R. A. (2010). Knowledge sharing: A review and directions for future research. *Human Resource Management Review*, 20(2):115–131.
- Wright, J. (2013). Punishment strategies in repeated games: Evidence from experimental markets. *Games and Economic Behavior*, 82:91–102.
- Zhao, X. and Xue, L. (2012). Competitive target advertising and consumer data sharing. *Journal of Management Information Systems*, 29(3):189–222.