Industrial data sharing and data readiness: a law and economics perspective

Nathalie Jorzik[1,2] · Paula Johanna Kirchhof[3] · Frank Mueller-Langer[3,4]

Abstract

We study the incentives and welfare properties of industrial data sharing taking into account the data (economy) readiness of companies. We differentiate between two regulatory settings. First, there is no compulsion for companies to provide data. Companies, which also use the data for other corporate purposes, decide whether to share their data voluntarily. Second, there is a regulatory requirement on the minimum amount of data to be shared by the data provider. We assume that data sharing affects the data provider's value of the data. The magnitude and sign of this effect have an impact on the optimal investment level of data generation and overall welfare in the different cases under study. Our results suggest that the implementation of a data-sharing policy has ambiguous welfare properties. It has positive welfare properties if (a) the data receiving firm does not pay too much for the data, (b) the data receiving firm benefits enough from the data provider's data generating effort, and (c) the intensified competition due to data sharing is not too harmful to the data provider. In contrast, it will always have negative welfare properties if the data provider's minimum amount of data to be shared under the policy is prohibitively high such that no data is created in the first place. Our results also suggest that a positive effect of data sharing on the data-generating company's value of the data and its data economy readiness positively affect the incentives to share data. Finally, we find that data sharing under a data-sharing policy leads to a lower data quality if the data economy readiness of the data-generating company is too low.

✉ Frank Mueller-Langer
frank.mueller-langer@unibw.de

[1] Alfred Weber Institute for Economics, University of Heidelberg, Heidelberg, Germany

[2] BMW Group, Munich, Germany

[3] Department of Business Administration, University of the Bundeswehr Munich, Neubiberg, Germany

[4] Max Planck Institute for Innovation & Competition, Munich, Germany

1 Introduction

Data is a non-rival good in the sense that one person's use of the data does not reduce or diminish another person's use. Multiple users can access the data at the same time (Varian, 2019). In addition, data can be used repeatedly without impacting the data quality or running the risk that the supply of data will be depleted. While the volume of data generated by humans and machines has increased exponentially in recent years, the data are often generated and stored by just a few large companies (Martens, 2023; Martens & Duch-Brown, 2020).

The EU Commission estimated that approx. 80% of machine-generated data in the data economy remains unused to date arguing that the use of these data may increase economic growth and innovation (European Commission, 2022). The call for encouraging Business-to-Business data sharing has intensified recently (Richter & Slowinski, 2019).

However, several concerns have been raised about the implementation of data sharing and data markets, especially when it includes personal data (Spiekermann & Acquisti, 2015). The question whether personal data could be seen as "property" is subject of a fierce debate in law, economics, and policy (Acquisti & Varian, 2005; Goldfarb & Tucker, 2011; Samuelson, 2000; Schwartz, 2003; Spiekermann & Novotny, 2015).[1]

Nevertheless, the EU Commission's Data Act provides a legal framework of data access and use, according to which companies should in the future make their data more easily available to consumers, public authorities and third parties.[2]

In recent years, companies that generate and store huge amounts of data have formed cooperations with other companies to build up data ecosystem in order to share data among themselves. Handling and sharing data is a key component of these ecosystems. Following Cattaneo et al. (2020) and Glennon et al. (2022), they are one of the key factors which may help to drive the European Data Market towards higher growth and push the contribution of the data economy to the EU GDP up to 4 percent.

---

[1] Another challenge associated with personal data markets relates to the difficulty to capture the value of personal data (Acquisti et al., 2013; Berthold & Böhme, 2010). Technical security issues associated with markets for personal data, e.g., data encryption and protection from personal data breaches, as well as social and ethical issues, e.g., the "propertization of the human being" and the difficulty to define what constitutes a "digital identity", are also fiercely debated in science and policy (Sackmann et al., 2006; Spiekermann & Acquisti, 2015).

[2] For recent analyses of data access from a competition law perspective, see Kerber (2019a, 2019b), and Martens and Mueller-Langer (2020). For recent works on data markets from an economics perspective, see Bonatti and Bergemann (2012), Bergemann et al. (2022), Acemoglu et al. (2022), and Koutroumpis et al. (2020).

However, companies may be reluctant to share their data for various reasons. Besides strategic motives for access denial such as the protection of the competitive advantage associated with the data, they might not engage in data sharing because they fear a loss of control over their data when it is re-used by third parties. Hence, the challenge is to implement instruments that facilitate access to data which is normally not being made available. Low trust, conflicting economic incentives and technological obstacles impede the full realisation of the potential of data-driven innovation (European Commission, 2022; Duch-Brown et al., 2017).

According to European Commission (2022), the Data Act includes measures which will enable consumers to access the data of their connected devices and use it for aftermarket and value-added services, e.g., predictive maintenance.

However, the Data Act, in its current form, is subject to a fierce debate in science, policy and industry (Drexl et al., 2022; Kerber, 2022; Metzger & Schweitzer, 2022; VDA, 2022).

As for the requirements for data-generating companies regarding the provision of data, Article 4(1) of the Data Act indicates the following:

> "Where data cannot be directly accessed by the user from the product, the data holder shall make available to the user the data generated by its use of a product or related service without undue delay, free of charge and, where applicable, continuously and in real-time."

This may require a high level of efforts on the part of the data-generating companies. However, companies may only have sufficiently high incentives to exert high efforts to create and prepare the data for re-use if they can monetize the data downstream. If monetizing the data downstream is difficult or even legally prohibited, this approach could eventually inhibit innovation, as companies may launch fewer products that generate data. Based on this, companies may be forced to exert high levels of effort to make data available even if they are of negligibly low relevance and have low potential for downstream monetization. In this respect, the aforementioned requirements may lead to distorted incentives to invest in data-driven products and services in the first place. The underlying trade-off is at the core of our analysis.[3]

In order to explore the trade-off between the societal benefits of industrial data sharing and the cost incurred by the data-generating industry, we set-up a simple model of industrial data sharing. We consider two players, i.e., a data-generating manufacturer and a data-reusing company, and study two data-sharing policies. Under No Data-Sharing Policy, the manufacturer can freely choose whether to share the data. In contrast, under Data-Sharing Policy, it is mandatory for the manufacturer to share a minimum amount of data. In both settings the manufacturer chooses the effort to generate and prepare the data. We assume that data sharing affects the data provider's value of the data. Our main results are as follows.

---

[3] Other fiercely debated issues in science and policy are that the Data Act in its current form may threaten the protection of know-how and the confidentiality of industry trade secrets and that a reverse flow of data from service providers to product manufacturers is not foreseen in the Data Act.

First, we find that the implementation of a data-sharing policy has ambiguous welfare properties. It has positive welfare properties if the data receiving firm does not pay too much for the data and benefits enough from the data provider's data generating effort while the intensified competition due to data sharing is not too harmful to the data provider. In contrast, it will always have negative welfare properties if the data provider's minimum amount of data to be shared under the policy is prohibitively high such that no data is created in the first place. Next, our results suggest that a positive effect of data sharing on the data-generating company's value of the data and its data economy readiness positively affect the incentives to share data. We also find conditions on the imposed minimum amount of data to be shared such that the manufacturer will not create any data under a data-sharing policy. Finally, we find that data sharing under a data-sharing policy leads to a lower data quality if the data economy readiness of the data-generating company is too low.

The remainder of the paper is organised as follows. Section 2 provides an overview of the literature on industrial data sharing and data economy readiness. In Sect. 3, we set up a simple model of industrial data sharing. Section 4 presents our results. In Sect. 5, we discuss our results, and Sect. 6 concludes.

2 Literature

The European Commission published a Proposal "for a Regulation on harmonized rules on fair access to and use of data" (Data Act) in February 2022 (European Commission, 2022). The conclusions of the 2020 council meetings on the Data Act highlighted the importance for ready-available high-quality data (European Council, 2020). The Data Act builds upon this by stating the following in Article 114(1): "The same dataset may potentially be used and reused for a variety of purposes and to an unlimited degree, without any loss in its quality or quantity." Hence, the Data Act aims for consistent data quality and quantity between the creation of the data to the purchase and the re-use of data. Based on this, the question whether the quality and quantity of data remains unchanged in a regulated data market is at the core of the present study.

The Data Act stresses possible benefits of industrial data sharing to increase innovation through lowering the barriers of entry and decreasing data monopolies by stating the following in Article 25: "The data tends to remain under the control of the manufacturers, making it difficult for users to obtain value from the data generated by the equipment they purchase or lease. Consequently, there is limited potential for innovative smaller businesses to offer data-based solutions in a competitive manner and for a diverse data economy in Europe." Hence, the Data Act acknowledges that data has a direct influence on the business performance and therefore the revenue of companies (see also Richter & Slowinski (2019)). Generally, the Data Act extends the "Data Governance Act" of 2020 by suggesting possible 'horizontal' rights and mandatory provisions on Business-to-Business (B2B) data exchange. Notably, in the setting suggested by the Data Act, the commercialisation of data is not fully specified (Metzger & Schweitzer, 2022). Hence, it is still uncertain how

businesses will be motivated to share their data or whether they are possibly forced to share it. Therefore, in our model outlined in Sect. 3, we provide an analysis of different scenarios.

2.1 Industrial data sharing

As Prüfer and Schottmüller (2021) and Zuboff (2019) suggest, companies collect data on their users which in return helps them to adapt their products to the users' preferences. This, in turn, increases sales which leads to decreasing marginal costs of innovation. This may lead to a first-mover advantage with a tendency to monopolization where the increasing data-driven indirect network is unachievable for new entrants which ultimately may hinder their innovativeness (Prüfer & Schottmüller, 2021).

Intervening in the market by forcing data-generating companies to share their data to governments is a fiercely debated issue in science and policy (see, for instance, Martens & Duch-Brown (2020)). This idea gained support at the beginning of the Covid-19 pandemic when public voices demanded that Google should allow governments to access location data to decrease infestation (Cukier et al., 2022).

Currently, market failures hinder market-sharing opportunities due to data monopolies, high transaction costs of data sharing, possible risks associated with data sharing, and low incentives of companies to share their data (Martens et al., 2020).[4] Notably, industrial data sharing may be a way to address the issue of data monopolization.

Overall, one may distinguish the market for data into a semi-public market or a private market where data is provided in exchange for a payment. In our model, the first scenario under study is similar to a private, unregulated data market where, under certain conditions, industrial data will not be provided in the absence of a data-sharing policy. An important aspect of the model is that the value and quantity of the data for re-use purposes depend on the effort to generate and prepare it in the first place. In addition, they depend on the data-generating company's data economy readiness.[5] However, under the second scenario under study, our model assumes the presence of a data-sharing policy which requires companies to provide a minimum amount of data. Based on this, important economic issues such as the quality of data and the cost to supply it in a usable, readily available format are addressed in our model.

The economics literature on industrial data sharing is still relatively scarce. Notable exceptions are Koutroumpis et al. (2020), Martens et al. (2020), and Martens and Duch-Brown (2020).[6] Our paper is related to Mueller-Langer and Andreoli-Versbach

---

[4] See also van Panhuis et al. (2014) for an overview of the impediments for data sharing in public health.

[5] Data economy readiness is often also referred to as data readiness. Henceforth, we use the two terms interchangeably.

[6] For interesting parallels, see also the recent literature on academic data sharing (Alter & Gonzalez, 2019; Fecher et al., 2015) See also Graef and Prüfer (2021) who set-up a model of industrial data sharing and appropriate governance frameworks. They suggest that monopolization, which is currently insufficiently regulated through the GDPR, must be mitigated. Based on this, they propose a governance framework for data sharing which complements the GDPR and the Data Act. In particular, they put forward a 'Data Pool' in which data-generating companies pool their anonymized data.

(2018) which studies the effect of mandatory data disclosure on the data-sharing incentives and welfare in science. Our paper differs from Mueller-Langer and Andreoli-Versbach (2018) in several important aspects. First, we consider industrial data sharing where–while data may be shared even if the price of data equals zero–we also allow for monetary incentives that may spur data sharing. In contrast, Mueller-Langer and Andreoli-Versbach (2018) consider a scenario where there are no monetary incentives for data sharing. Second, the model set-up in Mueller-Langer and Andreoli-Versbach (2018) is specific to research and publication of empirical papers in science. In their model, data disclosure only occurs when the respective article is published in a journal with a mandatory data-disclosure policy, i.e., data disclosure would never occur without journal publication. In contrast, in our model voluntary data sharing may occur in the absence of a data-sharing policy if data sharing has a sufficiently large positive effect on the data producer's value of the data. Third, we account for the data readiness of the data producer and the data-receiving firm while Mueller-Langer and Andreoli-Versbach (2018) do not account for this aspect.

Overall, recent literature on the barriers for industrial data sharing suggests that the likelihood of data sharing decreases in the cost of data production (Azkan et al., 2022; Johnson et al., 2017; Arnaut et al., 2018; Frontier Economics, 2021; Godel et al., 2022; Martens et al., 2020; OECD, 2019). In a similar fashion, prior literature on the barriers for data sharing in science suggests that the cost of data production negatively affect the likelihood that data-producing researchers share their data (Costello, 2009; Feigenbaum & Levy, 1993; Kim & Stanton, 2013; McCullough et al., 2006).[7]

Based on the aforementioned arguments and findings, our analysis explores the incentives of actors to share their data while accounting for effort cost, data quantity, data quality and data value.

2.2 Data readiness

Büchel and Engels (2022) assess a company's data readiness by measuring their data storage capacities, data management and processing level as well as their usage of the data.[8] In their survey, they discover that 71% of the participating companies have a low level of data readiness. As part of their questionnaire they also examine the participating companies' current data-sharing activities. They report a low level of data-sharing activity correlating with a low level of data readiness. Moreover, the process of preparing the data for sharing can also depend on the technical attributes of data, e.g., the data's volume, velocity, variety, viscosity, and veracity (Olama

---

[7] Andreoli-Versbach and Mueller-Langer (2014) provide evidence that, once the cost of data production and preparation are sunk due to the publication of an empirical article in a scientific journal with a mandatory data-disclosure policy, the likelihood of voluntary data sharing - via public data archives - increases.

[8] In contrast, in the medical field, the term 'data readiness' refers to the idea that data is directly usable for a specific application (Douthit et al., 2021).

et al., 2014). These dimensions influence a company's data readiness. For example, if the data variety of a given company is high, e.g., the data is from many different sources, it might need more effort to combine these data into a single bundle, which can then be shared. Another framework to assess the complexity of data readiness is proposed by Castelijns et al. (2020). They introduce "bands" (C, B, A, AA, and AAA) that stand for the level onto which a company's data readiness can be classified.[9] Overall, these findings suggest that a company's data readiness is quantifiable and affecting its data-sharing abilities. In our model, we distinguish between the data readiness of the data producer and the one of the data-receiving firm.

2.3 Data cost, data quality, and data value

The introduction of the GDPR required companies to change their data gathering, processing and usage strategies. In some cases, the costs for the usage of some specific data types which were majorly relevant before the implementation of the GDPR became obsolete due to its high costs of anonymization or processing according to the GDPR rules (Barati et al., 2020). Hence, the costs of data to prepare to adhere to regulations is an important driver of the decision to fetch and save the data. In the following, we will address the cost of preparing data and its effect on the quantity, quality and value of the data in more detail.

Recent works suggest that legal concerns and data security issues, e.g., data leakage to third parties and data abuse, pose a significant hurdle to the willingness of companies to share data (Azkan et al., 2022; Demary et al., 2019; Krotova et al., 2020; Röhl et al., 2021; Yaodong & Shuai, 2022).[10] This directly influences the costs of preparing the data. For instance, when the risk of data abuse increases, the costs to encrypt the data will *ceteris paribus* also increase. Errors and lengthened processes to conduct analyses are a typical sign of lower data quality. Accordingly, the quality of data is directly related to the effort spent in creating and cleaning it (Batini et al., 2009). For example, data standardization, linkage and schema integration are tasks which are preventive costs to ensure that higher data quality is achieved (Eppler & Helfert, 2004).

Existing data-sharing pools in the medical field show that, if data with low quality is shared, incremental costs arise when such data is combined with higher-quality data for AI models (Skripcak et al., 2014). Hence, the data quality not only directly relates to the effort to gather and clean the data prior to sharing but also impacts the costs of the data receiver when including and re-using it.

The nominal value of data is determined by many different internal and external factors (see, for instance, Duch-Brown et al. (2017)). Data value is not only based on the demand for the data (individual value) but also on its effect on the business area or industry (economic value). The different kind or type of data ultimately influences the value of the data set. This may be particularly true if we assume that the initial value of the data is equal to the cost of production of the data plus a factor *x* for profit. For example, location data is a relatively readily available source for some data-driven

---

[9] Their assessment includes parts from both Olama et al. (2014) and Büchel and Engels (2022) such as quantity and management evaluation. It also includes a welfare and legal analysis.

[10] See also Chen et al. (2019).

companies as their products or services are location dependent such that the location of a user is asked for, e.g., delivery services, marketplaces, or dating apps. Therefore, location data will have a relatively low production cost in this case. For such companies, the value of location data—if only measured by its costs of production—might be low as their efforts to gather this type of data are relatively low. Other companies, however, might struggle to gather location data as their products or services are location independent or only require a country location (e.g., e-commerce, retail, or news). Research has shown that location data is a key driver to generate high returns on marketing campaigns for online retail, also known as geo-marketing (Andrews et al., 2016). Hence, for those companies, the value of location data is relatively high leading to a high-demand of location data which in return would increase the profit of the company which can gather location data at a low cost.

Finally, using a data-sharing platform in farming as an example, Wysel et al. (2021) show that managing data is a key part of the cost of data while it also positively influences the value of data. Distinguishing between efficient and sustainable data-value creation, their analysis suggests that data-value creation depends on the effort to prepare the data. Applying this concept to our model, this means that companies which opt for efficient data-value creation might opt to save costs when preparing data for purchase. For instance, they may reduce data-cleaning actions which, in turn, reduces the value of the data. In contrast, companies which aim to create sustainable data value may employ more advanced methods to clean the data. This, in turn, will ease the employment of the data once acquired, i.e., it will reduce the cost of data re-use.

Overall, prior literature suggests that the cost of data production and data sharing vary substantially across industries (Azkan et al., 2022; Godel et al., 2022; Grody et al., 2006). On the one hand, in the biomedical industry, the costs of creating clinical trial data, scans, experimental and laboratory data are very high, e.g., high upfront investments in the technical equipment that scientists need to create and prepare such data. Rockhold et al. (2016) suggest that the cost and required resources for data sharing are major barriers for the sharing of clinical trial data. Based on this, sharing of clinical trial data rarely occurs as overall costs are disproportionately high. On the other hand, there are industries where the cost of data production are relatively low because they are a by-product of another production process, with virtually no additional production cost for the data producer (Duch-Brown et al., 2017; Hugenholtz, 2016). For instance, car manufacturers generate data about the emotional responses of their drivers with smart car systems with relative low effort (Swan, 2015). Another example is eBay's market price data being a by-product of its auction activities. In this case, there is arguably no reason for the data producers to stop the data production process even if they do not generate any profits from granting data access to third parties. Notably, eBay shares a large variety of market price data and other sales data via several application programming interfaces (APIs).[11]

---

[11] See https://developer.ebay.com/develop/apis/restful-apis, last accessed 13 July 2023. See also eBay's Terapeak tool that gives users broad access to sales data (https://www.ebay.com/help/selling/selling-tools/terapeak-research?id=4853#section1, last accessed 13 July 2023.

Finally, prior works also suggest that the value spillovers of data and overall effects of mandatory data sharing may vary across industries (de Vries et al., 2023; Teeters et al., 2008). For instance, in the computational neuroscience industry, computational models are developed that integrate experimental data in order to explore the brain function. The experimental data is typically acquired by highly specialized experimenters. However, theorists that are highly specialized in analysis methods often do not have access to the experimental data leading to a sub-optimal level of exploitation of existing data sets (Teeters et al., 2008). Based on this, there are high value spillovers of data and data sharing as only the combination of datasets from several sources may promote the productivity of the industry by allowing new insights, meta analyses, and a better match of skills and resources. Connected to this, mandatory data sharing is likely to have a positive effect on the overall productivity of the neuroscience industry (Teeters et al., 2008). In contrast, industrial data sharing may have a negative effect in highly competitive industries where proprietary and highly sensitive information play a crucial role, e.g., the banking, pharmaceutical or healthcare industries (He et al., 2023; Ke & Sudhir, 2023; Stach et al., 2022). In this case, data sharing may reduce the competitive advantage associated with the data and lead to safety and privacy issues (Godel et al., 2022; Martens et al., 2020; OECD, 2019).

Based on the aforementioned arguments and findings, we assume in our model that the effort to gather and clean data positively affects (a) the quantity of the data, (b) the data-creating company's value of the data, and (c) the data-using company's value of the shared data.

3 A simple model of industrial data sharing

Following Mueller-Langer and Andreoli-Versbach (2018), we set-up a simple model of industrial data sharing. We analyze the optimal effort choices of a manufacturer, M, to create and prepare data and to share the data with another company, C, who may re-use the data for her own business purposes. In our three-stage model, $t = 0, 1, 2$, the incentives to create, prepare and share data depend on two factors. First, it depends on the impact that data sharing has on the manufacturer's utility. Second, M's incentives to share data will depend on her data readiness. The idea behind this aspect is that if M's data readiness is low, she incurs higher cost of creating, preparing and sharing the data to C. For instance, the lower M's data readiness, the higher will be her cost of setting up and implementing a data-sharing ecosystem, and the lower will *ceteris paribus* be her incentives to share the data.

We study two data-sharing policies. Under No Data-Sharing Policy (henceforth, *NP*), M can freely choose whether to share the data. In contrast, under Data-Sharing Policy (henceforth, *P*), it is mandatory for M to share a certain amount of data with C. The motivating example for the data-sharing policy under study is the EU Data Act, which establishes access requirements for the industry to provide data. We explore under which conditions this may increase the effort cost of the data-generating company thereby reducing the incentives to invest in data creation.

Under both policies, M chooses the effort to create and prepare the data in stage 0, e_0. For simplicity, we assume that the quality of the data is equal to the effort to create it. The cost of data creation and preparation incurred by M are given by $c_M = \frac{1}{2}e_o^2$. M's value of the data depends positively on e_0. Provided that the data is shared, C's value of the data also depends positively on e_0. In $t = 1$, M decides whether to share the data. In $t = 2$, provided that the data is shared, C generates utility from re-using the shared data. If the data is shared, C exerts effort e_c. The cost of using the data incurred by C are given by $c_C = \frac{1}{2}e_c^2$. Let v_M be M's value of the data. We assume that v_M depends positively on e_0. The intuition behind this assumption is as follows. The higher the effort in data creation and preparation, the higher will be the quality and quantity of the data and thus M's value of the data. In addition, $M's$ value of the data is directly affected by her data readiness. The underlying idea is that companies with a higher data readiness can *ceteris paribus* store, manage and process their data more efficiently. M's data readiness is given by α_M with $\alpha_M > 0$.

In our model, M chooses the effort to create the data taking the market price of the data, p, as given.[12] The quantity of the data is given by $x(e_0) = \alpha_M \cdot e_0$. It depends positively on both M's effort to create and prepare the data as well as on M's data readiness.

As Fig. 1 illustrates, there are four possible scenarios.

Under *NP*, M may choose to (1) not share the data with C in $t = 1$ (henceforth, indicated by NP^{no_share}) and (2) share it with C in $t = 1$ (NP^{share}). In the latter case, C generates utility by re-using the data in $t = 2$, while he does not generate any utility in $t = 2$ in the former case.

Under *P*, there are regulatory requirements for M on the minimum amount of data to be shared, as given by $\bar{x}$. Arguably, if $\bar{x}$ is prohibitively high, the manufacturer may find it optimal not to invest in the creation and preparation of the data under a data-sharing policy. Based on this, we consider two possible scenarios under *P*. (3) M chooses an effort in $t = 0$, which leads to data sharing in $t = 1$ (P^{share}). (4) Due to a prohibitively high minimum quantity of data to be shared, $\bar{x}$, M finds it optimal to not exert any effort to create and prepare the data in $t = 0$ such that data sharing does not take place in $t = 1$ (P^{no_share}).

We assume that $M's$ value of the data changes when it is shared with C. The underlying idea is that data sharing may have countervailing negative and positive effects on M depending on the *exogenously given* competitive environment under which M and C operate in the product market, which in turn will depend on the type of $M's$ data. In our model, we account for these possibly countervailing effects as follows. β, with $0 < \beta$, measures the extent to which $M's$ value of the data changes when it is shared with C.

If $0 < \beta < 1$, the negative effects of data sharing on $M's$ value of the data more than outweigh its positive effects. In this case, by sharing the data, the data-generating company may reduce its competitive advantage associated with the data by too

---

[12] See also Sect. 5.2 where we explore the case that, following Article 4(1) of the Data Act, M is required to provide the data to C free of charge, i.e., we assume that $\bar{p} = 0$.
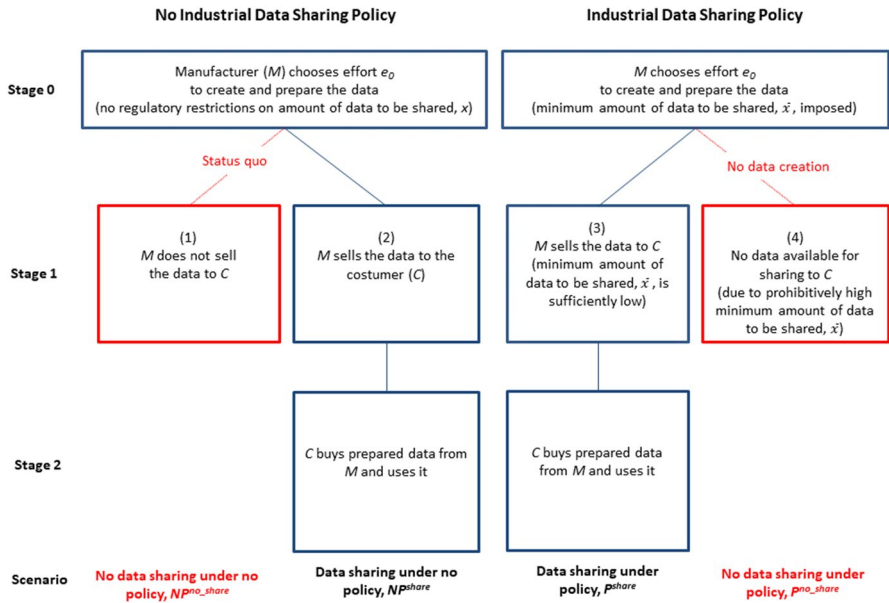
Fig. 1 Timing of interactions

much (Godel et al., 2022).[13] In addition, one may argue that data sharing not only allows competing companies to re-use the shared data, adding value to their own data, but also to expose the data-sharing company's data gathering methods or the type of data they collect. Based on this, data sharing may have a negative overall effect on the data-sharing company's value of the data. In the aforementioned cases, β would be rather low and, in the extreme, tend toward zero.

In contrast, our model also captures the case that $\beta > 1$, i.e., the competitive environment and the type of $M's$ data is such that the positive effects of data sharing on $M's$ value of the data dominate. This may arguably be the case if the data is re-used as a form to create a new data set which is functionally different. In this case, the data-using downstream innovator C would not compete against the creator of the original data set M but instead would complement the data (Duch-Brown et al., 2017). Hence, there would be no competition due to the re-use of data and $M's$ incentives would be fundamentally different as compared to a situation where the entrance of a new competitor in the market may lead to economic downturns. In addition, data sharing may have a positive effect on M's reputation (Singh et al., 2020; Thomas & Leiponen, 2016). In this case, β may be large and, possibly, larger than one.

---

[13] In their study on data-driven mergers, de Cornière and Taylor (2022) argue in a similar fashion. They suggest that the incentives for data-driven mergers will be higher the larger is the role that data plays in competition.

In the following Sect. 3.1, we consider the two cases under *NP* as illustrated by Fig. 1. Then, we consider the two cases under *P*.

3.1 No industrial data sharing policy

We first consider the no-data-sharing case under *NP*. Then, we explore data sharing under *NP*.

3.1.1 No data sharing

In this scenario, *M* chooses effort e_0 to create and prepare the data but does not share it with *C*. *M*'s maximization problem is given by:

$$u_{M,NP^{no\_share}} = v_{M,NP^{no\_share}} - \frac{1}{2}e_o^2, \tag{1}$$

where *M*'s value of the data is given by $v_{M,NP^{no\_share}} = \alpha_M \cdot e_0$, with $\alpha_M > 0$. The second term on the right-hand side of Equation (1) indicates *M*'s cost of data creation and preparation, $c_M = \frac{1}{2}e_o^2$.

As no data are shared under this scenario, *C*'s utility equals zero.

3.1.2 Data sharing

In the second scenario, *M* chooses effort e_0 to create and prepare the data and shares the data with *C*. *M*'s maximization problem is given by:

$$u_{M,NP^{share}} = v_{M,NP^{share}} + p \cdot x - \frac{1}{2}e_o^2, \tag{2}$$

where *M*'s value of the data is given by $v_{M,NP^{share}} = \beta \cdot \alpha_M \cdot e_0$ with $\beta > 0$ and $\alpha_M > 0$. It increases in the extent to which the positive effect of data sharing dominates, as given by an increasing β, and in *M*'s data readiness. The second term on the right-hand side of Equation (2) is the product of the price at which the data is shared with *C*, *p*, and the quantity of the data $x(e_0) = \alpha_M \cdot e_0$.

C's utility depends on *M*'s effort to create and prepare data, its own data readiness, as given by $\alpha_C > 0$, its own effort, e_C, and parameter κ, $0 < \kappa < 1$. κ reflects the fact that *C* only benefits partially from *M*'s data generating effort, e_0, as according to the EU Data Act not all effort that *M* exerts has to be shared with *C*.[14] Based on this, *C*'s utility is given by:

$$u_{C,NP^{share}} = v_{C,NP^{share}} - p \cdot x - \frac{1}{2}e_C^2, \tag{3}$$

---

[14] According to the EU Data Act, data-producing firms only have to share raw data but not processed data nor inferred insights. In addition, they are allowed to withhold the data in order to protect trade secrets or avoid irreparable economic damage. We thank an anonymous referee for this suggestion.

Proposition 1 (Data Sharing under No Data-Sharing Policy) (*I*) *If $\beta \geq 1$, the data are always shared under no data sharing policy.* (*II*) *If $\beta < 1$, (i) the data is shared if $p \geq 1 - \beta$, and (ii) it is not shared if $p < 1 - \beta$.*

The intuition behind these results is as follows. For values $\beta \geq 1$, the positive effect of data sharing dominates and the manufacturer is willing to share her data at any price, i.e., she will share the data even if the price is zero. For values $\beta < 1$, the manufacturer's willingness to share her data depends on the relation between β and p. If β is very low, the manufacturer will be willing to share her data only if p is sufficiently high. If, however, β is close to 1, the manufacturer will be willing to share her data even if p is very low.

4.2 Data sharing policy

We explore the question under which conditions on the imposed minimum amount of shared data, $\bar{x}$, M will not create any data under P.

Proposition 2 (No Data Sharing under Data-Sharing Policy) (*I*) *Under a data-sharing policy, the data is not shared if, for the minimum required quantity of shared data, $\bar{x}$, the following condition holds: $\bar{x} > \alpha_M^2 (\beta + p)$.* (*II*) *The data is shared if $\bar{x} \leq \alpha_M^2 (\beta + p)$.*

The intuition behind these results is as follows. The higher $\bar{x}$, the lower is *ceteris paribus* the likelihood that the manufacturer is willing to share her data. The lower $M's$ data readiness, as given by a lower α_M, the higher is the likelihood that the condition specified in (I) holds. In addition, the larger the extent to which the negative effects of data sharing on $M's$ value of the data outweigh its positive effects, as given by a lower β, the higher is the likelihood that the condition specified in (I) holds. Note that, in (II), data sharing may occur even if $p = 0$, i.e., this is the case when α_M and β are sufficiently high.

4.3 Welfare effects of the transition from *NP* to *P*

We explore the question under which conditions the transition from *NP* to *P* leads to a deterioration in the quality of the shared data as given by the effort to create it.

Proposition 3 (Lower Data Quality under Transition from *NP* to *P*) (*I*) *If $\beta \geq 1$ the transition from NP to P leads to lower data quality if $\alpha_M^2 (\beta + p) < \bar{x}$.* (*II*) *If $0 < \beta < 1$ and i) $p \geq 1 - \beta$ the transition from NP to P leads to a lower data quality if $\alpha_M^2 (\beta + p) < \bar{x}$; and ii) if $p < 1 - \beta$ the transition from NP to P always leads to a lower data quality.* (*III*) *In all other cases, the transition from NP to P has no effect on the quality of the data.*

where $v_{C,NP^{share}} = \kappa \cdot e_0 \cdot e_C \cdot \alpha_C$.

3.2 Industrial data sharing policy

Under a data-sharing policy, M chooses effort e_0 to create and prepare the data and shares the data with C under the condition that $x(e_0) = \alpha_M \cdot e_0 \geq \bar{x}$. M's maximization problem is given by:

$$u_{M,P^{share}} = v_{M,P^{share}} + p \cdot x - \frac{1}{2}e_0^2, \text{ under the condition that } x(e_0) \geq \bar{x}, \qquad (4)$$

where M's value of the data is given by $v_{M,P^{share}} = \beta \cdot e_0 \cdot \alpha_M$.

C's utility is given by:

$$u_{C,P^{share}} = v_{C,P^{share}} - p \cdot x - \frac{1}{2}e_C^2, \qquad (5)$$

where $v_{C,P^{share}} = \kappa \cdot e_0 \cdot e_C \cdot \alpha_C$.

Here, the main aspect of the model is the following. If the imposed minimum amount of data to be shared $\bar{x}$ is sufficiently small, M will choose an effort e_0, which leads to data sharing data under P (P^{share}). To illustrate, consider the extreme case that $\bar{x} = 0$. In this case, M's optimization problem under P^{share} is equivalent to the one under NP^{share}. However, if $\bar{x}$ is prohibitively high, the negative effects of shared data on M may be so high that M omits data sharing by not creating any data (P^{no_share}). To illustrate, consider the extreme case that $\bar{x}$ tends to infinity. In this case, M will find it optimal not to create any data. In Proposition 2, we derive the conditions on $\bar{x}$ under which there is no data sharing under P.

4 Results

Table 1 summarizes the results for M's optimal effort to generate and prepare the data, M's utility, the quantity of the data, and overall welfare, as given by the sum of M's and C's utility, for each regime illustrated in Fig. 1. We calculate the welfare level for each scenario to analyze whether there are parameter constellations in which data sharing is not optimal for the producer but optimal for society. Note that, as Table 1 illustrates, the overall welfare depends on the data readiness of both the data producer and the data-receiving firm if the data is shared under NP and P. The proofs behind these results are straightforward and are, thus, omitted here.[15]

4.1 No data sharing policy

We explore the question under which conditions data is shared under no data-sharing policy.

---

[15] All proofs are available upon request.

Table 1 Optimal effort to create data, utility, quantity, and overall welfare

Regime	Effort	Utility of manufacturer	Quantity	Welfare
NP^{no_share}	$e_0^* = \alpha_M$	$u_M^* = \frac{1}{2} \cdot \alpha_M$	$x^* = \alpha_M^2$	$W^* = \frac{1}{2} \cdot \alpha_M^2$
NP^{share}	$e_0^* = \alpha_M(\beta + p)$	$u_M^* = \frac{1}{2}(\alpha_M(\beta + p))^2$	$x^* = \alpha_M^2(\beta + p)$	$W^* = (\frac{\kappa^2\alpha_C^2+1}{2})\beta^2\alpha_M^2 + (\frac{\kappa^2\alpha_C^2-1}{2})p^2\alpha_M^2 + p\cdot\beta\cdot\alpha_M^2\cdot\alpha_C^2\cdot\kappa^2$
P^{share}	$e_0^* = \alpha_M(\beta + p)$	$u_M^* = \frac{1}{2}(\alpha_M(\beta + p))^2$	$x^* = \alpha_M^2(\beta + p)$	$W^* = (\frac{\kappa^2\alpha_C^2+1}{2})\beta^2\alpha_M^2 + (\frac{\kappa^2\alpha_C^2-1}{2})p^2\alpha_M^2 + p\cdot\beta\cdot\alpha_M^2\cdot\alpha_C^2\cdot\kappa^2$
P^{no_share}	$e_0^* = 0$	$u_M^* = 0$	$x^* = 0$	$W^* = 0$

The intuition behind these results is as follows. (I) If $\beta \geq 1$ the manufacturer will always share her data under NP with $e_{0,NP^{share}}^{*} = \alpha_M(\beta + p)$. The transition from NP to P will lead to lower data quality if the manufacturer decides not to share any data with $e_{0,P^{noshare}}^{*} = 0$. This is the case if $\bar{x}$ is sufficiently high with $\bar{x} > \alpha_M^2(\beta + p)$.

(II) If $0 < \beta < 1$ and i) $p \geq 1 - \beta$ the manufacturer always shares her data under NP with $e_{0,NP^{share}}^{*} = \alpha_M(\beta + p)$. This level of data quality can only be achieved under P if $\bar{x} \leq \alpha_M^2(\beta + p)$, i.e., data sharing takes place under P. In contrast, if $\bar{x} > \alpha_M^2(\beta + p)$ the transition from NP to P lowers the data quality as no data are generated by the manufacturer. (ii) If $p < 1 - \beta$ the manufacturer does not share her data under NP with $e_{0,NP^{noshare}}^{*} = \alpha_M$. If $\bar{x} > \alpha_M^2(\beta + p)$ no data are generated under P with $e_{0,P^{noshare}}^{*} = 0$. Thus, $e_{0,NP^{noshare}}^{*} > e_{0,P^{noshare}}^{*}$. In contrast, if $\bar{x} \leq \alpha_M^2(\beta + p)$ the manufacturer shares her data under P with $e_{0,P^{share}}^{*} = \alpha_M(\beta + p) < \alpha_M = e_{0,NP^{noshare}}^{*}$ if $p < 1 - \beta$.

(III) If the data is shared under both NP and P, it will have the same quality. Finally, note that the transition from NP to P can never lead to a higher data quality. The intuition behind this result is the following. From Table 1 we can see that the transition from NP to P may only have a positive effect on e_0^* if M does not share the data under NP while she shares the data under P. This is the case if $e_{0,NP^{noshare}}^{*} < e_{0,P^{share}}^{*}$ with $\alpha_M < \alpha_M(\beta + p)$ from which follows $1 < \beta + p$. This, however, contradicts the condition for no data sharing under NP in Proposition 1(ii) that $1 > \beta + p$.

Now, we analyze the overall welfare effects of the transition from NP to P.

Proposition 4 (Ambiguous Welfare Effects of Transition from NP to P) *(I) If $\beta \geq 1$ the transition from NP to P leads to a lower welfare level if $\alpha_M^2(\beta + p) < \bar{x}$. (II) If $0 < \beta < 1$ and i) $p \geq 1 - \beta$ the transition from NP to P leads to a lower welfare level if $\alpha_M^2(\beta + p) < \bar{x}$. ii) If $p < 1 - \beta$ the transition from NP to P has negative effects on welfare if $\alpha_M^2(\beta + p) < \bar{x}$. If $\alpha_M^2(\beta + p) \geq \bar{x}$, the transition from NP to P has positive welfare effects for the following parameter constellations: $\alpha_C^2 \cdot \kappa^2 > \frac{1 + p^2 - \beta^2}{(\beta + p)^2}$. For $\alpha_C^2 \cdot \kappa^2 < \frac{1 + p^2 - \beta^2}{(\beta + p)^2}$, the transition from NP to P has negative welfare effects. (III) In all other cases, the transition from NP to P has no effect on welfare.*

The intuition behind these results is as follows. (I) If $\beta \geq 1$ the manufacturer will always share her data under NP. The transition from NP to P leads to a decrease in welfare if the manufacturer decides not to share any data under P with $W_{P^{noshare}}^{*} = 0$. This is the case if the regulatory requirements on the data quantity $\bar{x}$ are prohibitively high with $\bar{x} > \alpha_M^2(\beta + p)$. (II) If $0 < \beta < 1$ and i) $p \geq 1 - \beta$ the manufacturer always shares her data under NP. This welfare level can only be achieved under P if $\bar{x} \leq \alpha_M^2(\beta + p)$, i.e., data sharing takes place under P. In contrast, if $\bar{x} > \alpha_M^2(\beta + p)$ the transition from NP to P lowers the welfare level as the manufacturer decides not to share her data anymore. (ii) If $p < 1 - \beta$ the manufacturer does not share her data under NP with $W_{NP^{noshare}}^{*} = \frac{1}{2} \cdot \alpha_M$. If $\bar{x} > \alpha_M^2(\beta + p)$ no data are generated under P with $W_{P^{noshare}}^{*} = 0$. In this case, the welfare effect of

a transition from *NP* to *P* is negative. In contrast, if $\bar{x} \leq \alpha_M^2(\beta + p)$ the manufacturer shares her data under *P* and the overall impact on welfare depends on the relation between $W^*_{NP^{noshare}}$ and $W^*_{P^{share}}$. We obtain that, for certain parameter constellations of α_C, κ, p and β, overall welfare increases, i.e., $W^*_{NP^{noshare}} < W^*_{P^{share}}$. In Fig. 2a and b, we display different parameter constellations for which the overall welfare effect is positive.

(III) If the data is shared under both *NP* and *P*, the transition from *NP* to *P* will have no effect on welfare.

Finally, note that, in contrast to Proposition 3 where the transition from *NP* to *P* can never lead to an increase in data quality, here the transition can lead to an increase in welfare. The three-dimensional Fig. 2a and b illustrate this aspect. In both figures, the price at which the data is sold, *p*, is given by the *x*-axis. The *y*-axis shows the extent to which data sharing has a negative effect on *M*'s value of the data, β. The extent to which *C* (partially) benefits from *M*'s data generating effort, κ, is given by the *z*-axis.

In Fig. 2a we consider the case where the data readiness of *C* is relatively low, i.e., we set $\alpha_C = 0.75$. In Fig. 2a, the orange area is given by parameter constellations where both welfare levels are equal, i.e., $W^*_{NP^{noshare}} = W^*_{P^{share}}$.[16] Fig. 2a shows that, *above* the orange area, there exist parameter constellations for β, κ and *p* for which the overall welfare effect of the transition from *NP* to *P* is positive. The grey area represents all parameter constellations where κ reaches its maximum value, i.e., it tends toward one. For all parameter constellations *below* the orange and grey areas, the welfare effect is negative. We observe the following three aspects from the grey area in Fig. 2a. First, data sharing can never have positive welfare properties if β tends toward zero, i.e., data sharing has the largest possible negative overall effect on *M*'s value of the data. Second, data sharing can never have positive welfare properties if the price of the data, *p*, tends toward one. Third, if κ tends toward zero, such that *C* hardly benefits from *M*'s data generating effort, for almost all parameter constellations of β and *p*, data sharing can never have positive welfare effects. Based on this, for data sharing to have positive welfare properties the following conditions have to be satisfied: *p* is sufficiently low while β and κ are sufficiently close to one. Stated differently, data sharing has positive welfare properties if *C* does not pay too much for the data and benefits enough from *M*'s data generating effort while *M*'s value of the data does not decrease by too much if the data is shared.

In Fig. 2b, we consider the case where the data readiness of *C* is relatively high, i.e., we set $\alpha_C = 1.5$. Comparing the orange areas in Fig. 2a and b, we can see that the orange area and the set of parameter constellations for β, κ and *p above* the orange area, for which the welfare effect is positive, are larger in Fig. 2b. This is due to the relatively higher data readiness of *C* in Fig. 2b as compared to Fig. 2a.

---

[16] This is the case when $\alpha_C^2 \cdot \kappa^2 = \frac{1+p^2-\beta^2}{(\beta+p)^2}$ with $0 < \beta < 1$ and $p < 1 - \beta$ (see Proposition 4II).
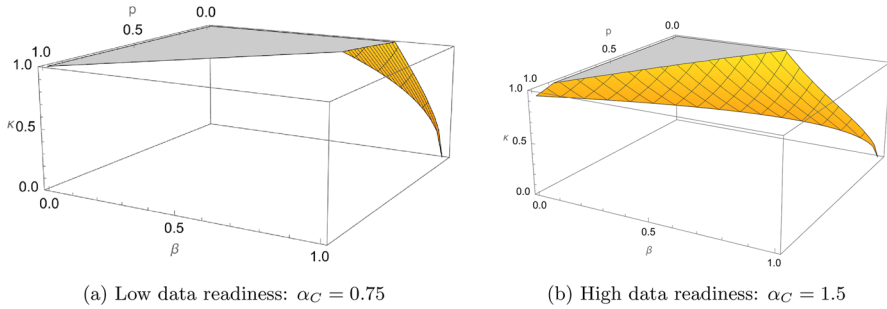
(a) Low data readiness: $\alpha_C = 0.75$ (b) High data readiness: $\alpha_C = 1.5$

Fig. 2 Positive welfare properties of the transition from *NP* to *P*. *Notes*: **a** and **b** illustrate parameter constellations where the overall welfare effect of a transition from *NP* to *P* is positive. In both figures, the orange area is given by parameter constellations where overall welfare is the same under both *NP* and *P*. For parameter constellations *above* the orange area the transition from *NP* to *P* has positive welfare properties. The orange area and the set of parameter constellations for β, κ and p *above* the orange area are larger in **b** than in **a**. This is due to the relatively higher data readiness of *C* in **b**

5 Discussion

5.1 Assumptions

In our model, data sharing may occur under *NP* and *P* even if $p = 0$. However, we also consider the case where data can be traded at a positive price. In this regard, we follow a recent strand of economics literature on data markets (Bonatti & Bergemann, 2012; Bergemann et al., 2022; Acemoglu et al., 2022; Koutroumpis et al., 2020).[17] Here, the underlying assumption is that the data can be directly exchanged between *M* and *C*. However, one may also think of a scenario where a third party intermediates between *M* and *C*.[18] For instance, data-sharing platforms may provide the technical infrastructure for the exchange of data between multiple parties. From an economic perspective, their key function is to facilitate data sharing by lowering transaction costs through combining different data sources and matching users and suppliers (see Richter and Slowinski (2019)). For instance, in the context of access to digital car data, a "neutral server" architecture is discussed whereby data storage and data processing will be provided by a third-party data intermediary (Martens & Mueller-Langer, 2020). An example for a car-data platform is Otonomo, which provides real-time and historical traffic data to its customers in exchange for a payment.[19] While the analysis of data intermediaries is beyond the scope of the present paper, it is an interesting idea for further research.

---

[17] See also Drexl (2017) for a proposal for designing competitive markets for industrial data from a competition law perspective.

[18] In the context of personal consumer data, see Ichihashi (2021) for a model of competition between data intermediaries.

[19] See https://otonomo.io/. Another example is the car-data platform Caruso (https://www.caruso-dataplace.com). See also Wysel et al. (2021) for an analysis of data-sharing platforms for agricultural data.

We also assume that $M's$ incentives to share data depend on her data economy readiness as given by α_M. Recent survey evidence from 1,002 companies from industrial and industrial-related service sectors in Germany provides empirical support for this assumption (Büchel & Engels, 2022). Büchel and Engels (2022) suggests that there is a positive correlation between a company's data economy readiness and the role that data sharing plays for the company. This study also provides empirical evidence that 71% of the surveyed companies are not data-economy ready and that for 73% of them data sharing plays no role. Based on this, one may argue that the α_M and α_C parameters in our model may be rather low.

Finally, we assume that, under P, there is a regulatory requirement on the minimum quantity of data to be shared by the manufacturer. One can think of alternative ways how such a data-sharing policy may be implemented. We outline a possible extension of the model with respect to the price of data in Sect. 5.2.

5.2 Extensions

5.2.1 Free-of-charge data sharing

An interesting extension of the model might be to explore additional ways how the data-sharing policy may be implemented. For instance, consider the case that the regulator imposes a ceiling on the price of data, $\bar{p}$. While a full analysis of this case is beyond the scope of the present paper, we consider the main effects of $\bar{p} = 0$ in our model. This assumption is based on Article 4(1) of the Data Act which requires the data-generating company to make the data available to the data user "free of charge". For $\bar{p} = 0$, we obtain the optimal effort to create data, utility of M, quantity and welfare as given by Table 2.

Comparing the results reported in Tables 1 and 2, we can see that $M's$ optimal effort to create the data, $M's$ utility, the data quantity, and overall welfare under NP^{share} and P^{share} would *ceteris paribus* decrease if $\bar{p} = 0$ and $\beta < 1$. That is, if the negative effect of data sharing on $M's$ value of the data more than outweighs its positive effect, one may argue that distorted incentives for data sharing and negative welfare properties of the transition from NP to P are *ceteris paribus* more likely to occur when a "free of charge" data-provision requirement is imposed on M alongside a minimum amount of data to be shared. In contrast, if the positive effect of data sharing on the value of the data dominates, i.e., $\beta > 1$, there will be positive data-sharing incentives for M under NP and also under P even if $\bar{p} = 0$.

Table 2 Optimal effort to create data, utility, quantity, and welfare if $\bar{p} = 0$

Regime	Effort	Utility of manufacturer	Quantity	Welfare
NP^{no_share}	$e_0^* = \alpha_M$	$u_M^* = \frac{1}{2} \cdot \alpha_M$	$x^* = \alpha_M{}^2$	$W^* = \frac{1}{2} \cdot \alpha_M{}^2$
NP^{share}	$e_0^* = \alpha_M \cdot \beta$	$u_M^* = \frac{1}{2}(\alpha_M \cdot \beta)^2$	$x^* = \alpha_M^2 \cdot \beta$	$W^* = (\frac{\kappa^2 \alpha_C^2 + 1}{2})\beta^2 \alpha_M^2$
P^{share}	$e_0^* = \alpha_M \cdot \beta$	$u_M^* = \frac{1}{2}(\alpha_M \cdot \beta)^2$	$x^* = \alpha_M^2 \cdot \beta$	$W^* = (\frac{\kappa^2 \alpha_C^2 + 1}{2})\beta^2 \alpha_M^2$
P^{no_share}	$e_0^* = 0$	$u_M^* = 0$	$x^* = 0$	$W^* = 0$

5.2.2 Ideas for further research

An interesting idea for further research would be to explore the possibility of joint ventures driven by data sharing. In parallel to data sharing in science, where data sharing may lead to joint projects between the data-sharing researcher and the data-reusing researcher, industrial data sharing may lead to data-driven joint ventures. While these joint ventures may have positive overall welfare properties due to increased overall innovation, they may also increase the data-sharing company's value of the data. In the case of data-driven joint ventures, key questions are the degree of protection of databases and contractual solutions for data sharing and their implications on the re-use of the data (Duch-Brown et al., 2017; Fries & Scheufen, 2019).[20]

In our model, we assume that the competitive environment under which M and C operate in the product market is exogenously given. In the model, this aspect is captured by the effect of data sharing on $M's$ value of the data, as measured by β. This simplifying assumption allows us to focus the present paper on studying the data-producing firms' incentives to share industrial data with service providers using their data while keeping the model tractable. However, it is an interesting avenue for further research to more explicitly explore the interaction between M and C in the product market.

Another interesting avenue for further research is to endogenize the data readiness of M and C. In the present model, we treat α_M and α_C as exogenous parameters in order to be able to focus on the endogenous data-investment efforts and their impact on the welfare properties of a transition from NP to P. For follow-up work, we suggest to analyze the data readiness of the data producer and the data-receiving firm as endogenous variables that the firms can invest in. Relatedly, our results suggest that the data readiness of firms may affect the overall benefits of a regulatory policy of data sharing. This in turn raises the question if and how the data readiness of firms may be promoted alongside the implementation of a regulatory policy of data sharing. In this respect, recent works on the obstacles of data sharing in disruptive technologies, e.g., autonomous systems, suggest that nudging, nodality and treasury policy tools such as industry guidelines in building robust data infrastructures, regulatory data-sharing sandboxes, or public-private collaborations may promote the data readiness of companies thereby decreasing technical, economic, and political barriers for data sharing (Tan et al., 2023; Tan & Crompvoets, 2022).

Another interesting idea for further research might be to address the specific role of the regulator or government in the context of industrial data sharing. In the present paper, we consider business-to-business (B2B) data sharing following Martens et al. (2020). Follow-up work may extend the model to business-to-government (B2G) data sharing following Martens and Duch-Brown (2020). While B2G data sharing is beyond the scope of the present paper, it is interesting to outline the new

---

[20] Arguably, there are parallels between the theoretical reasoning behind data sharing and licensing of patents. For instance, Richter and Slowinski (2019) argue that data sharing mirrors the theory of licensing of patents where intellectual property rights influence the market and its innovation capabilities.

players and decisions in this new model. In this case, one may introduce a third player in the model, i.e., the government G, in addition to M and C. Then, one may model the minimum amount of data to be shared, $\bar{x}$, as an endogenous decision of the government to maximize overall welfare.

6 Conclusion

We set up a simple model describing the incentives of a company to invest in data creation, use data and share it with another company. We consider two regulatory scenarios taking into account the data economy readiness of the data-generating company and its competitive advantage associated with the data. First, under NP ("no data-sharing policy"), the data-creating company can freely decide whether or not to share data voluntarily. Second, under P ("data-sharing policy"), there is a regulatory requirement on the minimum quantity of data to be shared by the manufacturer.

The implementation of a data-sharing policy may distort incentives in two ways. First, it may reduce the data-generating company's effort to create the data as data sharing reduces the competitive advantage associated with the data. Second, the imposed minimum quantity of data to be shared may be prohibitively high such that no data is created in the first place.

In terms of policy implications, our results suggest that the transition from NP to P is never beneficial to the data-generating manufacturer but may have positive welfare properties. Based on this, policymakers may implement mechanisms to increase the incentives to create and share data. In this respect, the policy recommendations derived from our analysis are in line with the Data Act's intention to maintain incentives for manufacturers to continue investing in high-quality data generation by covering their transfer-related costs and excluding the use of the shared data in direct competition with their products.

Finally, our model allows us to derive exact conditions for positive welfare properties of a transition from NP to P. We obtain positive welfare properties if (i) the minimum-quantity threshold under P is not too restrictive, i.e., $\alpha_M^2(\beta + p) \geq \bar{x}$, and (ii) if $\alpha_C^2 \cdot \kappa^2 > \frac{1 + p^2 - \beta^2}{(\beta + p)^2}$ holds. Condition (i) is *ceteris paribus* more likely to be met the higher is the data economy readiness of the data-generating company, α_M. In addition, the second condition is *ceteris paribus* more likely to be met the more C benefits from M's data generating effort, i.e., the higher is κ, and the higher is $C's$ own data economy readiness, α_C. Based on this, we argue that the estimation of α_M, α_C, and κ at the industry level (or company level) is an important empirical exercise in the context of industrial data sharing. However, while the estimation of α_M, α_C, and κ is beyond the scope of the present paper, we argue that, as companies in different industrial sectors are likely to have different levels of data readiness and κ, the welfare effects of a transition from NP to P may be different for different industrial sectors. This may eventually call into question the suitability of a "one-size-fits-all-industries" approach for data-sharing policies.

Declarations

Conflict of interest The authors declare that there are no financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

References

Acemoglu, Daron, et al. (2022). Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics, 14*(4), 218–56. https://doi.org/10.1257/mic.20200200

Acquisti, Alessandro, John, Leslie K., & Loewenstein, George. (2013). What is privacy worth? *The Journal of Legal Studies, 422*, 249–274.

Acquisti, Alessandro, & Varian, Hal. (2005). Conditioning prices on purchase history. *Marketing Science, 24*(3), 367–381.

Alter, George, & Gonzalez, Richard. (2019). Responsible practices for data sharing. *American Psychologist, 73*(2), 146–156. https://doi.org/10.1037/amp0000258

Andreoli-Versbach, Patrick, & Mueller-Langer, Frank. (2014). Open access to data: An ideal professed but not practised. *Research Policy, 43*(9), 1621–1633. https://doi.org/10.1016/j.respol.2014.04.008

Andrews, Michelle, et al. (2016). Mobile ad effectiveness: Hyper-contextual targeting with crowdedness. *Marketing Science, 35*(2), 218–233. https://doi.org/10.1287/mksc.2015.0905

Arnaut, Catarina et al. (2018). Study on data sharing between companies in Europe. Tech. rep. European Commission DG Communications Networks, Content and Technology, everis Benelux.

Azkan, Can et al. (2022). Incentives and economics of data sharing: Fields of action of cross-company data exchange and status quo of the German economy. Fraunhofer Institut für Software und Systemtechnik ISST, pp. 1–70.

Barati, Masoud, et al. (2020). GDPR compliance verification in internet of things. *IEEE Access, 8*, 119697–119709. https://doi.org/10.1109/ACCESS.2020.3005509

Batini, Carlo, et al. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys (CSUR), 41*(3), 1–52. https://doi.org/10.1145/1541880.1541883

Bergemann, Dirk, Bonatti, Alessandro, & Gan, Tan. (2022). The economics of social data. *The RAND Journal of Economics, 53*(2), 263–296. https://doi.org/10.1111/1756-2171.12407

Berthold, Stefan, & Böhme, Rainer. (2010). Valuating privacy with option pricing theory. In *Economics of information security and privacy* (pp. 187–209), Springer.

Bonatti, Alessandro, & Bergemann, Dirk. (2012). Markets for data. In *2012 Meeting Papers 538*. Society for Economic Dynamics.

Büchel, Jan, & Engels, Barbara. (2022). The importance of the data economy for Europe's digital strategic autonomy. In *Decoding EU Digital Strategic Autonomy Sectors, Issues, and Partners*. Ed. by

Gerard Pogorel, Antonios Nestoras, and Francesco Cappelletti. Brussels: European Liberal Forum, pp. 13–18. https://doi.org/10.53121/ELFTPS1.

Castelijns, Laurens A., Maas, Yuri, & Vanschoren, Joaquin. (2020). The ABC of Data: A Classifying Framework for Data Readiness. In Peggy Cellier & Kurt Driessens (Eds.), *Machine learning and knowledge discovery in databases* (pp. 3–16). Cham: Springer International Publishing.

Cattaneo, Gabriella et al. (2020). The European Data Market Monitoring Tool. D2.9 Final Study Report. European Commission, IDC, and The Lisbon Council.

Chen, Yuling, et al. (2019). FaDe: A blockchain-based fair data exchange scheme for big data sharing. *Future Internet, 11*(11), 225. https://doi.org/10.3390/fi11110225

Costello, Mark J. (2009). Motivating online publication of data. *BioScience, 59*(5), 418–427. https://doi.org/10.1525/bio.2009.59.5.9

Cukier, Kenneth, Mayer-Schoenberger, Viktor, & de Vericourt, Francis. (2022). Framers: Human Advantage in an Age of Technology and Turmoil. Penguin.

de Cornière, Alexandre, & Taylor, Greg. (2022). *Data and competition: A Simple framework with applications to mergers and market structure*. CEPR Press Discussion Paper No. 14446. Centre for Economic Policy Research (CEPR). https://cepr.org/publications/dp14446.

de Vries, Saskia E.J., Siegle, Joshua H., & Koch, Christof. (2023). Sharing neurophysiology data from the allen brain observatory. In: *eLife* 12. Ed. by Markus Meister et al., e85550.

Demary, Vera, et al. (2019). *Bereitschaft der deutschen Unternehmen für die Teilhabe an der Datenwirtschaft*. Gutachten, BMWi: Institut der Deutschen Wirtschaft, Köln.

Douthit, Brian J., et al. (2021). A conceptual framework of data readiness: The contextual intersection of quality, availability, interoperability, and provenance. *Applied Clinical Informatics, 12*(3), 675–685. https://doi.org/10.1055/s-0041-1732423

Drexl, Josef. (2017). Designing competitive markets for industrial data-Between propertisation and access. *JIPITEC, 8*(4), 257–292.

Drexl, Josef, et al. (2022). Position Statement of the Max Planck Institute for Innovation and Competition of 25 May 2022 on the Commission's Proposal of 23 February 2022 for a Regulation on Harmonised Rules on Fair Access to and Use of Data (Data Act). *Max Planck Institute for Innovation & Competition Research Paper, 22–05*. https://doi.org/10.2139/ssrn.4136484

Duch-Brown, Nestor, Martens, Bertin, & Mueller-Langer, Frank. (2017). The economics of ownership, access and trade in digital data. JRC Working Papers on Digital Economy 2017-01. Joint Research Centre, Seville. https://ideas.repec.org/p/ipt/decwpa/2017-01.html.

Eppler, Martin J., & Helfert, Markus. (2004). A classification and analysis of data quality costs. In *Proceedings of the 9th MIT Information Quality Conference* (pp. 311–325). http://www.alexandria.unisg.ch/54896/.

European Commission (2022). Proposal for a Regulation of the European Parliament and of the Council on Harmonised Rules on Fair Access to and Use of Data (Data Act). COM(2022) 68 final. European Commission.

European Council (2020). Special Meeting of the European Council (1 and 2 October 2020): Conclusions. EUCO 13/20, CO EUR 10, CONCL 6. European Council.

Fecher, Benedikt et al. (2015). A reputation economy: Results from an empirical survey on academic data sharing. Discussion Papers of DIW Berlin 1454. DIW Berlin, German Institute for Economic Research. https://doi.org/10.48550/arXiv.1503.00481.

Feigenbaum, Susan, & Levy, David M. (1993). The market for (ir)reproducible econometrics. *Social Epistemology, 7*(3), 215–232. https://doi.org/10.1080/02691729308578695

Fries, Martin, & Scheufen, Marc. (2019). Märkte für Maschinendaten: Eine rechtliche und rechtsökonomische Standortbestimmung. *MMR, 22*(11), 721–726.

Frontier Economics. (2021). Increasing access to data across the economy. Report prepared for the Department for Digital, Culture, Media and Sport. Frontier Economics.

Glennon, Mike et al. (2022). European data market study 2021–2023. D2.1 First Report on Facts and Figures. European Commission, IDC, and The Lisbon Council.

Godel, Moritz, Perkins, Ryan, & von Petersdorff, Clio. (2022). *Research into the Cost Considerations of Data Sharing*. London Economics: Final Report.

Goldfarb, Avi, & Tucker, Catherine E. (2011). Online advertising, behavioral targeting, and privacy. *Communications of the ACM, 54*(5), 25–27.

Graef, Inge, & Prüfer, Jens. (2021). Governance of data sharing: A law & economics proposal. *Research Policy, 50*(9), 104330. https://doi.org/10.1016/j.respol.2021.104330

Grody, Allan D., Harmantzis, Fotios, & Kaple, Gregory J. (2006). Operational risk and reference data: Exploring costs, capital requirements and risk mitigation. In *Journal of Operational Risk 1.3*. https://ssrn.com/abstract=849224.

He, Zhiguo, Huang, Jing, & Zhou, Jidong. (2023). Open banking: Credit market competition when borrowers own the data. *Journal of Financial Economics, 147*(2), 449–474.

Hugenholtz, P. Bernt. (2016). Something completely different: Europe's sui generis database right. In *The Internet and the Emerging Importance of New Forms of Intellectual Property*. Ed. by Susy Frankel and Daniel Gervais. Kluwer Law (pp. 205–222).

Ichihashi, Shota. (2021). Competing data intermediaries. *The RAND Journal of Economics, 52*(3), 515–537. https://doi.org/10.1111/1756-2171.12382

Johnson, Peter A., et al. (2017). The cost(s) of geospatial open data. *Transactions in GIS, 21*(3), 434–445.

Ke, T. Tony., & Sudhir, K. (2023). Privacy rights and data security: GDPR and personal data markets. *Management Science, 69*(8), 4389–4412.

Kerber, Wolfgang (2019a). Data governance in connected cars: The problem of access to in-vehicle data. In *JIPITEC* 9(3), pp. 310–331. http://nbn-resolving.de/urn:nbn:de:0009-29-48075

Kerber, Wolfgang. (2019). Data sharing in IOT ecosystems and competition law: The example of connected cars. *Journal of Competition Law & Economics, 15*(4), 381–426. https://doi.org/10.1093/joclec/nhz018

Kerber, Wolfgang. (2022). Governance of IoT data: Why the EU data act will not fulfill its objectives. *GRUR International*. https://doi.org/10.1093/grurint/ikac107

Kim, Youngseek, & Stanton, Jeffrey M. (2013). Institutional and individual influences on scientists' data sharing behaviors: A multilevel analysis. *Proceedings of the American Society for Information Science and Technology, 50*(1), 1–14. https://doi.org/10.1002/meet.14505001093

Koutroumpis, Pantelis, Leiponen, Aija, & Thomas, Llewellyn D. W. (2020). Markets for data. *Industrial and Corporate Change, 29*(3), 645–660. https://doi.org/10.1093/icc/dtaa002

Krotova, Alevtina, Mertens, Armin, & Scheufen, Marc. (2020). Open data and data sharing: An economic analysis. IW-Policy Paper 21. Institut der Deutschen Wirtschaft, Köln.

Martens, Bertin. (2023). *What should be done about Google's quasi-monopoly in search?* TILEC Discussion Paper: Mandatory Data Sharing Versus AI-Driven Technological Competition.

Martens, Bertin, & Duch-Brown, Nestor. (2020). The economics of business-to-government data sharing. JRC Working Papers on Digital Economy 2020-04. Joint Research Centre, Seville.

Martens, Bertin, & Mueller-Langer, Frank. (2020). Access to digital car data and competition in aftermarket maintenance services. *Journal of Competition Law & Economics, 16*(1), 116–141. https://doi.org/10.1093/joclec/nhaa005

Martens, Bertin, de Streel, Alexandre, et al. (2020). Business-to-business data sharing: An economic and legal analysis. JRC working papers on digital economy 2020-05. Joint Research Centre, Seville.

McCullough, B. D., McGeary, Kerry Anne, & Harrison, Teresa D. (2006). Lessons from the JMCB archive. *Journal of Money, Credit and Banking, 38*(4), 1093–1107.

Metzger, Axel, & Schweitzer, Heike. (2022). Shaping markets: A critical evaluation of the draft data act. *working paper* http://dx.doi.org/10.2139/ssrn.4222376

Mueller-Langer, Frank, & Andreoli-Versbach, Patrick. (2018). Open access to research data: Strategic delay and the ambiguous welfare effects of mandatory data disclosure. *Information Economics and Policy, 42*, 20–34. https://doi.org/10.1016/j.infoecopol.2017.05.004

OECD. (2019). *Enhancing access to and sharing of data: Reconciling risks and benefits for data re-use across societies*. Paris: OECD Publishing.

Olama, Mohammed M. et al. (2014). A qualitative readiness-requirements assessment model for enterprise big-data infrastructure investment. In *Next-generation analyst II*. Ed. by Barbara D. Broome, David L. Hall, and James Llinas. Vol. 9122. International Society for Optics and Photonics. SPIE, 91220E. https://doi.org/10.1117/12.2050605.

Prüfer, Jens, & Schottmüller, Christoph. (2021). Competing with big data. *The Journal of Industrial Economics, 69*(4), 967–1008. https://doi.org/10.1111/joie.12259

Richter, Heiko, & Slowinski, Peter R. (2019). The data sharing economy: On the emergence of new intermediaries. *International Review of Intellectual Property and Competition Law, 50*, 4–29.

Rockhold, Frank, Nisen, Perry, & Freeman, Andrew. (2016). Data sharing at a crossroads. *New England Journal of Medicine, 375*(12), 1115–1117. https://doi.org/10.1056/NEJMp1608086

Röhl, Klaus-Heiner, Bolwin, Lennart, & Hüttl, Paula. (2021). Datenwirtschaft in Deutschland. Wo stehen die Unternehmen in der Datennutzung und was sind ihre grössten Hemmnisse? Gutachten, BDI. Institut der Deutschen Wirtschaft, Köln.

Sackmann, Stefan, Strüker, Jens, & Accorsi, Rafael. (2006). Personalization in privacy-aware highly dynamic systems. *Communications of the ACM, 49*(9), 32–38.

Samuelson, Parmela. (2000). Privacy as intellectual property?, 52 Stanford Law Review 1125.

Schwartz, Paul M. (2003). Property, privacy, and personal data. In *Harvard Law Review* (vol. 117, p. 205).

Singh, Pranav Kumar, et al. (2020). Designing a Blockchain Based Framework for IoT Data Trade. In Siddharth Swarup (Ed.), *Innovations for community services* (pp. 295–308). Rautaray et al. Cham: Springer International Publishing.

Skripcak, Tomas, et al. (2014). Creating a data exchange strategy for radiotherapy research: Towards federated databases and anonymised public datasets. *Radiotherapy and Oncology, 113*(3), 303–309. https://doi.org/10.1016/j.radonc.2014.10.001

Spiekermann, Sarah, Acquisti, Alessandro, et al. (2015). The challenges of personal data markets and privacy. *Electronic Markets, 25*, 161–167.

Spiekermann, Sarah, & Novotny, Alexander. (2015). A vision for global privacy bridges: Technical and legal measures for international datamarkets. *Computer Law & Security Review, 31*(2), 181–200.

Stach, Christoph, et al. (2022). Protecting sensitive data in the information age: State of the art and future prospects. *Future Internet, 14*(11), 302.

Swan, Melanie. (2015). Connected car: Quantified self becomes quantified car. *Journal of Sensor and Actuator Networks, 4*(1), 2–29.

Tan, Evrim, & Crompvoets, Joep. (2022). A new era of digital governance. In *The New Digital Era Governance*. Wageningen Academic (pp. 13–49).

Tan, Si Ying, Taeihagh, Araz, & Pande, Devyani. (2023). Data sharing in disruptive technologies: Lessons from adoption of autonomous systems in Singapore. *Policy Design and Practice, 6*(1), 57–78.

Teeters, Jeffrey L., et al. (2008). Data sharing for computational neuroscience. *Neuroinformatics, 6*, 47–55.

Thomas, Llewellyn D. W., & Leiponen, Aija. (2016). Big data commercialization. *IEEE Engineering Management Review, 44*(2), 74–90. https://doi.org/10.1109/EMR.2016.2568798

van Panhuis, Willem G., et al. (2014). A systematic review of barriers to data sharing in public health. *BMC Public Health, 14*, 1–9. https://doi.org/10.1186/1471-2458-14-1144

Varian, Hal. (2019). Artificial Intelligence, Economics, and Industrial Organization. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda* (pp. 399–422). Chicago: University of Chicago Press.

VDA. (2022). Entwurf EU-Verordnung Data Act vom 23.02.2022: Regulierung und Privatautonomie für die Datenwirtschaft. Positionspapier. Verband der Automobilindustrie (VDA).

Wysel, Matthew, Baker, Derek, & Billingsley, William. (2021). Data sharing platforms: How value is created from agricultural data. *Agricultural Systems, 193*, 103241. https://doi.org/10.1016/j.agsy.2021.103241

Yaodong, Tao, & Shuai, Yang. (2022). Comparative study on data sovereignty guarantee technology. EasyChair Preprint 8965. IEEE.

Zuboff, Shoshana. (2019). Surveillance capitalism and the challenge of collective action. *New Labor Forum, 28*(1), 10–29. https://doi.org/10.1177/1095796018819461