Oliver Falck and *Johannes Koenen*

Resource "Data": Economic Benefits of Data Provision*

ABSTRACT

The importance of data for economic growth and development in Germany and Europe is undisputed. This study identifies various factors that may cause market failures for the good "data". As a consequence, markets alone may not lead to efficient allocation of, optimal access to and sufficient participation in data that is collected and stored. At present, different market-based, political and regulatory efforts are being undertaken in this context: State intervention such as a right to data access is pursued in parallel with private sector solutions such as data trusts or platforms. Given the complex issues underlying potential market failures, competition between different solutions is welcome.

Data has replaced oil as the world's most valuable resource over the past decade. In June 2008, according to the Financial Times Global 500 list, the four most valuable companies in the world were oil companies, whereas in mid-2018, the seven companies with the highest market capitalization were Internet and technology companies. Their business models are based, to a significant extent, on collecting, analyzing and using data. Whereas the success of oil companies relies on a resource that is finite and only available in certain places around the world, data is not subject to physical scarcity – on the contrary, current technological developments are leading to a rapid increase in both the amount of digital data available and its potential economic value. Collecting data has become much cheaper. Falling costs of digital sensors have accelerated the development of the Internet of Things (IoT) and proliferating smart devices generate streams of data. This process is likely to speed up further, as economic activity continues to shift to the Internet and increasingly complex digital devices are brought to market. For example, it is estimated that each autonomous vehicle produces three orders of magnitude (or more than 1,000 times) more usable data than the average Internet user (Schlosser 2018). Storing the data itself has become cheaper not only "technically" (falling prices for storage media), but also "organizationally": specialized market players such as cloud providers are providing solutions that exploit economies of scale on a grand scale and reduce the necessary initial investment for users (Carrière-Swallow and Haksar 2019). From the perspective of the collecting companies, this means that fixed investment costs are converted into variable costs. Beyond collection and storage, there are increasingly better possibilities for evaluating and analyzing collected data. Machine learning (ML) and artificial intelligence (AI) can identify underlying structures in existing data and generate forecasts or gain insights into user behavior. This may enable more efficient production, targeted advertising, automatic interaction with customers (via bots) or, in the near future, autonomous vehicles. Therefore, data is becoming increasingly valuable – and as a consequence the question of who controls data is drawing substantial attention.

German and European politics are increasingly focusing on access to and the free movement of data. The EU initiative for a Single European Data Space (European Commission 2017 and 2018) is a good example of this. From the German point of view, it is a particular concern to keep small and medium-sized enterprises competitive through access to data – this is expressed, among other things, in the key issues paper on SMEs by Economics Minister Altmaier (BMWi 2019).

After a brief description of the empirical role of data economics in the German and European economy, we examine which specific economic characteristics of "data" may cause market failures requiring

Oliver Falck

is the ifo Professor of Empirical Innovation Economics at the LMU Munich and Director of the ifo Center for Industrial Organization and New Technologies. Since 2012, he has also been Program Director of the CESifo Research Network.

Johannes Koenen

is the founder and CEO of ARC Econ GmbH, focused on applying insights and methods from economic research to policy advice and consulting. Previously, he has worked as a researcher on new technologies and industrial organization at the University of Mannheim, the University of Bonn and the ifo Institute.

* This research was financially supported by the Chamber of Industry and Commerce for Munich and Upper Bavaria.

regulatory intervention. We first analyze the incentives for data production and collection, followed by an examination of the extent to which (market) transactions and the exchange of data do or do not lead to efficient allocation of data. Only if and where there are market failures, for example in the face of externalities, a potential need for data regulation arises: This then raises questions on how to implement it in practice, which are discussed in the final section.

THE CONTRIBUTION OF DATA TO ECONOMIC PERFORMANCE IN EUROPE

Within the framework of the European Data Monitoring Tool, the European Commission has commissioned a study on how the data economy is developing in Europe (IDC 2019) based on a wide range of measures, including: the number of data-centric jobs and firms, the value of traded data-based services and products and the contribution of the data economy as a whole to the European gross national product. In 2018, 283,000 companies in Europe were classified as data providers – i.e., their main activity is to provide digital, data-based products and services. Compared to 2017, this number had risen by around 4.3%. The growth was even stronger for employees collecting, storing, managing, analyzing and visualizing data (an increase of 8.4% from 6.6 to 7.2 million). Therefore, 3.4% of all persons employed in the EU were working in the data economy. Growth appears to be limited by the lack of supply of data experts: in 2018, about 571,000 vacancies for data jobs in the EU could not be filled. Overall, the “data market” in the EU – i.e., the products and services based on the evaluation of data – was valued at EUR 71 billion – a substantial increase of 9.7% over the previous year. Taking multipliers (the data sector generates additional value for other industries) into account, the data economy generated around EUR 377 billion worth of output in the EU in 2018, or 2.6% of the overall economy. With a rate of 12%, the data sector is growing far faster than the overall economy on the continent (around 2%). This continued a lasting episode of extremely rapid growth (since 2014, the data economy in the EU has grown by around 50%). In addition to the lack of specialists mentioned above, a second central obstacle for further growth has been identified both by researchers and policymakers: The current regulatory environment requires reform, which will be a focus in the further course of this study.

CHARACTERISTICS OF THE “RESOURCE” DATA AND INCENTIVES FOR DATA COLLECTION

The focus of this study is data in digital form. We define data as digitally stored information that can be put into relation to other information and analyzed. For example, a temperature indication alone is not a data point; whereas, when it is combined with the time and place of measurement the information can be used for analysis or as an input into a service or product. Furthermore, it is valuable to know who (and, if applicable, with what type of instrument or sensor) collected the data point in order to assess the reliability of the observation (Koutroumpis et al. 2017).

Depending on the form in which data is available, the amount of effort required to analyze it differs substantially. One speaks of “unstructured” data, if the information is not organized in a database (or a comparable structure), but is, for example, distributed over various files and formats, or is available in a pure text form. Creating a structure (such as a database) in which data can be collected and organized and bringing it into a form that is conducive to analysis and evaluation requires effort and incurs substantial costs. Figure 1 illustrates this as one of the steps in the data value chain. Once data has been collected and structured (not necessarily by the same actor), it is passed on for analysis. At this point, the structured data is combined and enriched with further information, if necessary. Based on this dataset, systematic relationships in the data are examined using algorithms from the fields of AI and ML. The results of these analyses are in turn passed on to actors for whom they generate value. Based on the findings, e.g., advertising can then be tailored to the data subjects, or maintenance cycles of machines can be optimized.

Different settings are observed in practice. Each of these steps may be performed by the same actor – one example is Amazon: the platform observes user searches and purchasing behavior, evaluates it itself and finally places its own advertisements and recommendations (but also passes the information on to advertising partners). However, Figure 1 also suggests that, for example, a company specializing in ML approaches depends on access to data collected by others. Access to data is essential for SMEs and startups without data collection capacities, as well as in AI development, for example to “train” algorithms. How this access is implemented in practice depends to a large degree on the nature of the underlying data, as discussed below.

The Economic Value of Data

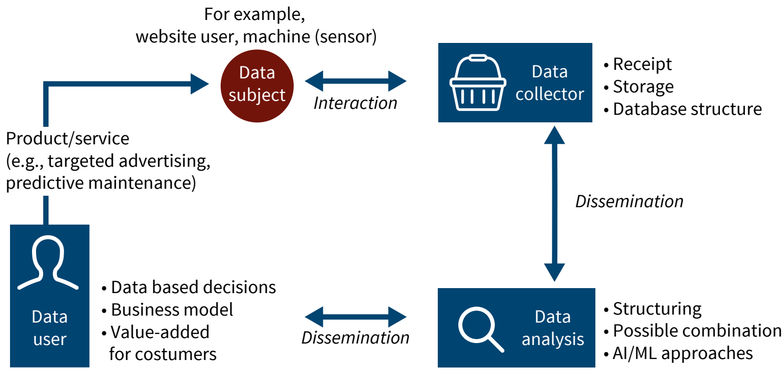
Conceptually, there are three central mechanisms for generating additional value from data and its analysis:

1. Through data-generated insights, business processes can be made more efficient and better decisions can be made (e.g., Brynjolfsson et al. 2011).
2. Integrating data enables the development of new products and services.
3. Data analysis potentially solves information problems and reduces information asymmetries, from which some market players can benefit.

Data may provide companies with more efficient organization or new "smart" products and services. However, point (3) above implies possible negative effects on some market participants. Lacking information can limit market efficiency: there may be less exchange than would be optimal and thus welfare is lost. However, a central result of information theory say that asymmetric information is associated with information rents for some actor(s). This includes some fundamental examples: If a retailer does not know a customer's exact willingness to pay, it cannot set prices in such a way that leaves no rents to each customer – even if the retailer has significant market power. The retailer therefore has an interest in learning its customers' willingness to pay as a pathway to obtaining a larger share of the consumer surplus through price differentiation or clever bundling of products. Conversely, if this happens, customers lose some of their surplus – they pay a higher price for the same product or service.

Figure 1

The Data Value-added Chain



Source: Authors' own compilation based on Li et al. (2019). © ifo Institute

Types of Data

Different classes of data can be distinguished according to their source of origin and content. We present a selection of the most important data types in Table 1.

In recent years, there has been considerable progress in the availability of **public data.** In addition to the EU portal mentioned above, comparable efforts are being made, for example, in the USA (data.gov), the UK (gov.uk), Austria (data.gv.at), as well as in various cities or by the London Transport Authority. The objective of these efforts is to create efficient access to public data via standardized interfaces (API). The aim is for companies to use this access to launch new or improved products and services.

With regard to the **machine-generated data,** some fundamental problems and conflicts of interest regarding the data are already apparent, which can probably be best illustrated using the example of data generated through motor vehicle operation (Kerber and Frank 2017). Considerable data streams are generated when a car is operated, especially in the context of navigation, safety systems and on-board diagnostic systems. The data subject here is the driver, and additional information is generated about his or her driving behavior. The right to collect and evaluate the vehicle's data is typically transferred to the vehicle manufacturer at the vehicle purchase (e.g., in the course of signing up for additional services). The manufacturer collects the resulting information in a structured form – in some cases in cooperation with cloud providers – and evaluates it in order to make forecasts about the wear and tear of parts, for example. However, there are other parties besides the manufacturer who are interested in accessing the data:

- Car parts manufacturers: their components provide part of the data (e.g., assistance systems).

Table 1

Selected Data Types

Data type	Description	Examples
Public, non-personal data	Data of public administration and authorities that are available in electronic form. There are international initiatives to make this data available to companies via standardized interfaces (API), such as the EU's open data portal.	Geographic maps, tendering databases, information on local and long-distance public transport, e.g., the Open Data initiative of Transport for London (Deloitte 2017).
Automatically generated data	Sensors and usage data of networked devices, machines and objects (IoT data).	Usage profiles of machines, data from the operation of motor vehicles, ambient temperature.
Data from internal IT systems of companies	Internal data required for the operation of the company, in particular from the areas of personnel, sales, logistics, customers, product quality and supplier management.	Personnel data (e.g., hiring, terminations by area), ERP data, CRM data, content of shared drives.
User and transaction data	Data resulting from the interaction of users with websites and platforms. This provides Information on completed transactions as well as the usage behavior and the length of stay on the website.	Logs and protocols of usage patterns and transactions, website cookies and other tracking.

Source: Authors' compilation.

By evaluating the information, the functionality of the parts can be improved.
- Workshop operators: these have an interest in access to wear information, for example, in order to be able to bring services to markets that compete with manufacturers' predictive maintenance offerings. Evaluating the vehicle's diagnostic systems is essential for maintenance and repair.
- Car insurance providers: aggregated driving profiles (e.g., by model) enable insurers to better assess the risks in the vehicle population. Access to individual driving profiles would make it possible to tailor individualized insurance offers.

In each of these cases, the interests of the data collector (manufacturer) and the other parties are not completely aligned. This suggests that manufacturers will tend to restrict data access for these actors (Kerber and Frank 2017). European legislators have considered this issue and the Regulation (EU) 2018/858 obligates manufacturers to grant independent actors access to large parts of the collected data (especially in the area of maintenance and diagnosis). This example is a clear indication that there could be generic problems with machine-generated data in similar constellations that are not solved by the market.

Internal data from company IT systems represent another interesting case. The issue here is not that other actors have a legitimate independent interest in the data, but rather that external specialists may be capable to initiate improvements in company processes by analyzing internal data or activating institutional knowledge that is partly lying idle on company hard drives. The potential of such analyses is reflected in the evaluation of the Munich start-up Celonis, which specializes in process analyses and has achieved a market value of more than one billion USD within less than ten years (Handelsblatt 2018).

Finally, the perhaps most discussed context of data collection is **user and transaction data**. An interesting special case here is the data on transactions of traders on online platforms, where the platform obtains and withholds information on the traders' own transactions. Here, the interests of the trader (building up an own, platform-independent customer base) sometimes collide with those of the platform (control over processes, primacy of the transaction on the platform, prevention of unwanted communication with customers). Depending on the type and origin of the data, different conflicts of interest and problems can arise.

The Economic Characteristics of Data

In order to analyze the reasons why regulation is potentially required in the context of data exchange and trade, it is also necessary to understand some specificities of data compared to other goods.

(1) Economies of Scale and Scope

Expensive infrastructure is needed to collect and analyze digital data: data centers with servers, storage media and software. Efficient data management and analysis require specialized skills and knowledge. Due to these factors, both economies of scale and economies of scope typically occur in connection with data.

- One speaks of economies of scale if the average costs incurred (e.g., per unit of stored data) decrease as the volume of data increases. Given that there are significant fixed costs when companies invest in data infrastructure and that the cost per additional unit of stored data is very low, economies of scale do exist (Duch-Brown et al. 2017).
- A related concept is economies of scope. Data collectors are able to process and analyze new data on related topics faster and at a lower cost, or to extract more value from them. The value of existing data on the road traffic situation in a city increases, for example, when information on load factors and delays in local traffic is added (Deloitte 2017). In this sense, different data can be complementary.

In practice, both mechanisms are further reinforced by the presence of **network effects.** Data collectors with a larger user-base generate a higher volume data stream, which for example enables faster progress in the development of AI and ML products, leading to better search results or user experience (Goldfarb and Trefler 2019). This in turn attracts new users, further enhancing the effect like a flywheel. At the same time, it should also be emphasized that economies of scale and scope are limited by the respective technological possibilities for storing and processing data (Varian 2014). Data sets can become too large and complex to be evaluated. Taken together, these effects thus provide, up to a certain point, significant economies of scale in data collection and analysis. They also explain why companies with a data-driven business model display such a "hunger for data" (Duch-Brown et al. 2017). In practice, this even has an impact on the structure of markets when acquisitions of companies are driven by data that targets have collected. On the other hand, due to the existence of scale and scope economies, it is potentially problematic from a welfare perspective if complementary data sets are kept separately by different actors (OECD 2019). In this context, one speaks of fragmentation, hoarding or silo formation.

(2) Non-rivalry and Limited Exclusivity

The analogy of data as resource is misleading in one important respect. Resources such as oil or gas are consumed in their use. The same data, on the other hand, can be analyzed and evaluated by any number of parties without affecting the information content and the knowledge gained (Carrière-Swallow and Haksar 2019). In the case of data, there is technological non-rivalry of use. However, a distinction must be made with regard to the incentives of data collectors. In many cases, the value of data results from the relative information advantage that users derive from it. This information advantage is automatically lost if all competitors have the same information at a given time. To put it bluntly: the first competitor invests in data-based target-group-marketing to increase the effectiveness of its advertising message. The second competitor invests to level out this advantage.

If data use leads to more efficient business processes, the widest possible use of this data would be desirable from a societal perspective, due to the non-rivalry of use. But here too, individual players have an incentive to hoard their data in silos in order to secure efficiency advantages over competitors. Private incentives thus tend to lead to too-low data sharing and too-little data exchange (London Economics 2019). At the same time, excluding other players from the use of data poses an organizational and technical challenge. In most cases, an interface to the outside world via the Internet is required to collect and analyze data, so in principle, access possibilities from outside also exist. In digital form, data can be duplicated and distributed at very low cost. To prevent this, i.e., to be able to actually exclude others from access and use, considerable investment in technical and organizational solutions is necessary. These efforts can be supported or hindered by the regulatory framework.

(3) Externalities

The collection of data may also involve significant negative externalities. In the context of personal data, the privacy of data subjects is affected. Furthermore, in the context of non-personal machine data, reducing information asymmetries can produce losers, for example, when manufacturers gain more precise information about the cost structure of their suppliers and adjust purchase prices accordingly.

On the other hand, there are possible positive externalities of data collection and data use. Up-to-date traffic data can reduce congestion and waiting times for all road users. In agriculture, data analyses can reduce the use of fertilizers and pesticides, and therefore contribute to improving the quality of groundwater (Wolfert et al. 2017). Reviews by hotel and restaurant guests help other consumers to make decisions.

Individual market participants disregard these external effects of data collection when making decisions – depending on the context, private incentives to collect data may therefore be too strong (driven by privacy and information rents) or too weak (in terms of reducing negative externalities).

(4) Data as an Intermediate Product or Raw Material

In its original form, data itself has little economic value. In order to generate value from raw data, it must be processed in several steps, some of which are time-consuming. These steps include designing a suitable database structure, collecting, evaluating and finally transferring it into a suitable business model for monetization. In this respect, data is actually comparable to a raw material or intermediate product (Jones and Tonetti 2018). In the data value chain, the end products are, for example, information on market segments, studies, analyses or services. It is at this stage of the value chain that a large part of the revenues of the data economy is generated, e.g., through ad auctions (Google) or the sale of ad space to customers with pre-selected characteristics (Facebook). The fact that these end products provide only limited insight into the underlying data makes it easier for integrated data companies – i.e., companies that cover all or several value creation stages – to protect their stored information from access (Duch-Brown et al. 2017). This means that those actors and companies that do not have approaches and skills for data analysis and use are, to a certain extent, lacking incentives to collect, structure and store their data, even if they have the potential to create considerable value added. Conversely, firms that are active at the various stages of the value chain accordingly have a reduced incentive to grant other actors (and thus potential competitors in the field of data analysis and evaluation) access to their collected data.

(5) Investment – Data as a By-product of Economic Activity

Finally, the question arises as to what extent the necessary (and, as presented, considerable) investment in data collection requires intervention by policy makers. In the area of innovative investment, it is well known that state support can provide targeted incentives to avoid underinvestment by private actors (Jaffe 1986). The stronger the (positive) externalities of data generation and the higher the incidence of free riders, the more likely it is that underinvestment may occur in the area of data economics (Duch-Brown et al. 2017). Consequently, in areas where negative externalities are more likely to occur (especially in personal data), even too high investments or too much collected data are to be expected. The extent to which state actors should influence the incentives to collect data thus depends strongly on the context and individual case. The need to differentiate here is further underscored by the fact that, in many cases, data is a by-product of the economic activities of companies. Examples of this are production or transaction data that are required for operations and may have to be stored for legal reasons. Once the neces-

Table 2

Causes of Market Failure in Data Transactions

Cause	Description	Effect
Asymmetric information – uncertain data quality	Potential buyers are subject to considerable uncertainty about the quality of data including: • Care in collecting • Consistency of formats etc. • Legality of the data ownership of the seller (e.g., compliance with legal regulations) • Completeness It is difficult to demonstrate data quality without providing access to the whole data set.	Higher complexity and thus higher transaction costs. Importance of relationship and trust between transaction partners, especially problematic for anonymous markets.
Non-transparent offer	There is no central market and no public directory for available data. Actors who would benefit from the analysis of existing data (sets) often have no knowledge of their existence. Intermediaries can reduce this problem, but they cause additional costs and are not yet relatively well established – and not available in all sectors. Since data is not (yet) a standardized product, it is difficult for buyers to compare the offers. Consequently, there are no »market prices« – according to market players, transaction prices are based on the (perceived) willingness of buyers to pay.	Advantageous and efficiency-enhancing transactions are potentially absent because buyers are unaware of the offer or sellers misjudge the willingness of buyers to pay.
Transaction costs	With regard to data transactions, quality standards and certification are still largely lacking. There are also no standardized »products« in terms of pre-defined data formats and database structures. In addition, the rights of the buyer in handling the data must be defined and his behavior must be checked, if necessary, with regard to: • Intended use • Right to combine with other data sets (risk of and de-anonymization of data subjects) • Passing on data, or analyses or services based on such data to third parties • Protection of data from unauthorized access after acquisition Compliance with regulatory requirements by purchasers (e.g., GDPR).	A lack of standardization makes detailed and therefore expensive contract drafting necessary. The resulting considerable transaction costs represent a market obstacle which particularly affects those smaller players who, for example, do not have a specialized legal department, to a greater extent.
Externalities	Analogous to the case of data collection, externalities on the part of the data acquirer can lead to the volume of data transactions being too low (positive externalities on the part of the acquirer, e.g., non-commercial actors), or too high (negative externalities, e.g., sending unwanted emails to acquired addresses).	The existence, magnitude and direction of the market failure depend on the context.
Market power, barriers to market entry	Market players behave strategically. In particular, vertically integrated companies (which collect data and also operate their own business models of exploitation) will often perceive buyers as potential or actual competitors; in such cases, there are strategic incentives to refrain from selling data in order to make it more difficult for others to enter the market.	Transactions and thus market access are made more difficult for players who do not have their own data sources.

Sources: Authors' compilation; Acquisti et al. (2016); London Economics (2019); Duch-Brown et al. (2017); Koutroumpis et al. (2017); and Carrière-Swallow and Haksar (2019).

sary data infrastructure is established, the marginal costs of data collection are extremely low (Farboodi and Veldkamp 2019).

DATA OWNERSHIP AND TRANSACTIONS

In the previous section, when examining incentives for data collection, the assumption was implicitly made that data collectors hold the rights to and have the control over the resulting data (-bases); this is very close to the legal reality in the United States, for example. Here we turn to the question of data ownership: what role do ownership rights, in particular the rights to use and exclude other actors, play in the context of data? Second, we analyze the barriers to data transactions or, in other words, how well a free market for data can function.

The Coase Theorem

The objective of data legislation and regulations should be the following: from a societal perspective, ownership of (or access to) data should ultimately be given to the actor who can generate the highest value (or benefit) from them. In economic theory, there is a simple solution to this: it is sufficient to define ownership rights to the object in a clear way. Via market transactions, the object should then be finally owned by the person with the highest willingness to pay – this is the basic logic of the Coase theorem (Coase 1960). In this case (which would be the second central insight), it is irrelevant from a welfare perspective to ask to whom the property is originally assigned. This is because the market ensures that it will end up with the "right" actor once a set of transactions has been completed. In order to achieve an efficient allocation through the market, two conditions must therefore be met: on the one hand, there must be clearly defined ownership rights to data; on the other hand, the market for data must function sufficiently well to enable and bring about the necessary transactions. If both conditions are met, there is no reason for the regulator to intervene in the market.

In the following section, we explain why the assumptions of the Coase theorem may not apply to data and in which areas there may be room for regulatory intervention. Furthermore, the original allocation of rights to data does play a central role in the distribution of rents between data subjects, collectors and users. Through the resulting investment incentives, this also affects, as discussed above, macroeconomic growth prospects and is thus of great importance for the economy as a whole (Acquisti et al. 2016).

Possible Causes of Market Failure

Which factors are potentially responsible for data markets failing or not developing at all (Koutroumpis et al. 2017)? Various properties of data and related market conditions can contribute to the fact that advantageous transactions of data are not possible and pure market mechanisms are thus not able to ensure an efficient allocation of data (London Economics 2019). In Table 2, we present an overview of the main causes of such market failures and their consequences in the context of data.

Each of the five factors analyzed above hampers the functioning of the market. The intensity of market failure, and hence the need for regulatory intervention, depends strongly on the individual context. In particular, a distinction should be made according to the size of the (potential) transaction parties. Transaction costs normally affect SMEs relatively stronger than large players with specialized legal departments – this applies both to contract costs and the costs of compliance with regulatory requirements (Koenen et al. 2018). Smaller market players are also less likely to employ specialized staff who observe the market environment in order to identify potential data sources. Thus, the intransparency of the offering tends to have a stronger impact on smaller companies as well. At the other end of the spectrum, there are cases where the market power of large, vertically integrated players "hoarding" data prevents potential competitors from entering the market.

Another point that deserves emphasis: the lack of standards and certificates for data transactions, together with the uncertain data quality in the run-up to the purchase, contributes to the fact that the relationship between buyer and seller plays an important role. If the actors trust each other (e.g., because of a grown business relationship or in expectation of further interactions in the future), the probability of misconduct decreases and the drafting of contracts becomes easier and cheaper. This in turn means that data transactions between larger, trusted parties are more likely to occur than the cases where smaller anonymous parties are involved (Duch-Brown 2017). This fact makes it more difficult for young, vertically non-integrated firms to enter the market, beyond strategic incentives of established players to implement entry barriers.

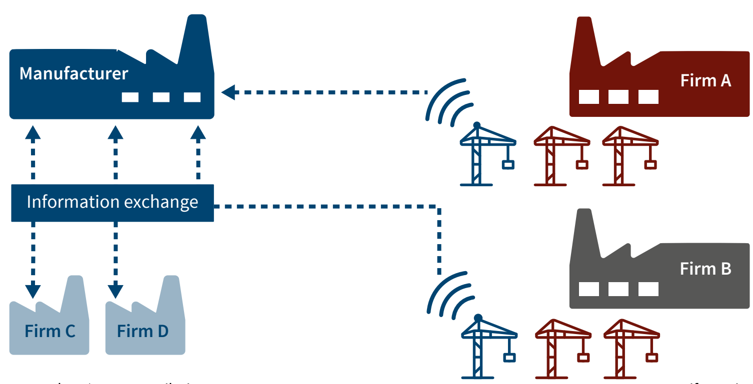
Case Study: Machine as a Service

The complexity of the factors (to be taken into account in data transactions) and the role of mutual trust are well illustrated by a case study which plays an increasingly important role in practice: the "Machine as a Service" (MaaS) model. For industry, and especially mechanical engineering in Germany, services play an increasingly central role in business models (Falck et al. 2019). This process is also known as "servitization". MaaS represents an important case: in this model, the customer no longer buys the machine from the manufacturer, but still receives and integrates it into his own production process in exactly the same way as he would with a purchased machine.

The difference is that instead of the fixed purchase price, the customer pays the manufacturer fees for the actual use of the machine. Instead of buying a compressor, for example, the customer purchases the "service" of compressed air from the manufacturer as demanded. Instead of the (high, one-off) fixed costs for the purchase of the machine, the user incurs variable costs that are completely based on use. In return, the manufacturer and provider of the service generates a more even payment flow. The provider usually guarantees the customer the complete functionality of the machine, i.e., he is also responsible for maintenance. The structure of the model is shown schematically in Figure 2.

Data plays a central role in the MaaS model: here the customer would have incentives to report lower-than-actual capacity utilization to the manufacturer. In addition, he might not take care of the

Figure 2
Machine as a Service - Schematic Diagram



Source: Authors' own compilation.

© ifo Institute

machine (e.g., let it get too hot), because the manufacturer is liable for breakdowns – a moral hazard for the customer. This problem can be solved by having the machine continuously send a stream of "real-time" data (related to usage, ambient temperatures, tool condition, etc.) to the manufacturer. Using this data, the manufacturer has a quasi-view of his customer's production process. Since he not only receives data from company A, but also from company B (and his other MaaS customers), he is able to detect possible misbehavior on the part of the customer with a high degree of probability, and is able to anticipate any maintenance work that may be necessary and therefore carry it out in a very efficient way. Without these data transfers, the MaaS model would not be viable. However, since the manufacturer's machine does not operate autonomously, but in interaction with other machines (e.g., compressed air as an energy source or turbines propelling an aircraft), the provider can gain insights into the customer's production process and its capacity utilization. The MaaS customer becomes "transparent" to a certain extent. The manufacturer, on the other hand, gains a "treasure trove" of data that can form the basis for new business models (Economist 2019).

For the MaaS model to work, it must be clarified what the manufacturer may use the collected data for. Customers must also be confident that the data infrastructure on the manufacturer's side is secured in such a way that sensitive information about their own business processes does not inadvertently fall into the wrong hands.

Consequences of Market Failures: Fragmentation, Data Silos

MaaS is a practical example of how data transfers in industry can lead to new business models. But what are the consequences if, due to market failure, shared access to data remains the exception or does not occur in some sectors? Companies that own and control data build silos where they store their data without giving access to active or potential competitors (Jones and Tonetti 2018). The ability to exclude others from using data blurs the boundaries between ownership and possession.

This fragmentation of data has two immediate negative effects. First, it is detrimental to competition, since firms in the data analysis field are effectively discouraged from entering the market. Second, it does not make efficient use of economies of scope, a key economic characteristic of data. The combination of complementary data sets can lead to increases in value, which can go unused if the necessary data sharing is not possible.

The fact that markets do not achieve efficient results derives from conditions of the Coase theorem being violated. Equally important, from a welfare point of view, it does matter to which party the original ownership rights to data are assigned. If data ownership is originally assigned to the data collectors, then this, combined with the data subjects' lack of market power and the existing network effects, is the basis for dominant market positions and high rents (Arrieta-Ibarra et al. 2018). If there are additional negative externalities of data use (e.g., with regard to private data), then an undesirable equilibrium results, in which large amounts of data are collected by multiple parallel players in an inefficient manner and, at the same time, too little data exchange takes place due to silo formation.

PROMOTING DATA SHARING AND ACCESS

International Laws and Regulation

Given the considerable economic importance of data, it is surprising to what extent the legal framework for data ownership is still unclear. In the status quo, which is particularly valid in the American lead market, data ownership is largely equivalent to ownership in terms of the use of data, unless personal rights of the individual make this difficult. Since, on the other hand, copyrights are not applicable to collected information, there is no legal possibility for the creators of databases to exclude others from using and duplicating them: they must rely on technical (copy protection, encryption) and organizational (secrecy) solutions. This increases the transaction costs for data, as described above.

The European legislators recognized relatively early that Europe is lagging behind the United States in terms of developing its data economy. The Database Directive (96/9/EC), adopted in 1996, was intended as an instrument to stimulate investment and the market in this area by defining ownership rights in databases. The Directive gives database creators two types of rights for a period of 15 years:

1. The structure of the database (but not the data it contains) is protected by copyright if it is "an intellectual creation of its author". This allows authors of protected databases to prevent other databases with identical structure from being made publicly available.
2. A new sui generis right is created which prohibits others from extracting or using substantial parts or all of the collected data. However, this is subject to the condition that the creation of the database requires "a substantial investment in qualitative or quantitative terms".

Official evaluations of the Directive conclude that the Directive has not had a significant impact and needs to be revised due to the fact that the creation of databases is increasingly automated.

In contrast, the European Data Protection Basic Regulation (GDPR, Regulation 2016/679) creates an effective legal framework for handling personal data, which has been in force since May 2018. The GDPR grants those data subjects who are natural persons a number of inalienable rights to their personal data, including in particular the rights of access, revocation, deletion, rectification, and a right to data transferability. Companies are bound by the principle of collecting only that data necessary for providing the service and using it only in the appropriate context. The GDPR is supplemented with regard to non-personal data by Regulation 2018/1807 on a framework for the free flow of non-personal data in the European Union. It aims at creating a European Data Area in which no additional barriers to data transfers exist caused by the national borders.

A detailed analysis of the GDPR is not the aim of this study. In summary, it can be said that, on the one hand, it substantially increases the regulatory requirements and the corresponding investment needs for companies that handle personal data. On the other hand, it establishes a clear legal framework for rights to data in the EU, which in the medium term can help to establish clear standards for the collection and transfer of data, which can reduce transaction costs.

Possible Solutions for Data Sharing

However, the problem of a lack of access to data, especially by small- and medium-sized enterprises, is not solved by the regulations mentioned above. In order to address this specific problem, various approaches are currently being put forward in the political and public debate, which we will now discuss.

(1) The Right to Data Access within the Value Chain

In the course of this study, we have identified various situations in which different parties had different, or even incompatible, interests with regard to data access. On the one hand, machine data can contribute to more efficient operation or better maintenance, yet it can, on the other hand, provide unwanted insights into the operator's production processes. In many cases, these problems can be solved through bilateral agreements, as shown in the MaaS example. This is more likely to succeed if the parties have a long-term business relationship and if they are larger players. By contrast, the problems appear to be more difficult to resolve if several rather small companies have a legitimate interest in access to data owned by another market player.

The German Ministry of Economics Paper on SMEs (BMWi 2019) indicates that politicians are considering a right to data access in value chains for SMEs. That such laws are within the realm of possibility is shown for example by EU Regulation 2018/858 discussed above, according to which other parties are also entitled to access automotive data. Through these approaches, legal compulsion is exerted to ensure the data "participation" of (smaller) market participants. In the case of automotive data, the data must be provided in a standardized "open" form, so that the necessary investments on the part of the data recipients remain relatively low. Therefore, one can anticipate that these approaches will be effective and actually allow access to the stored data.

However, it should be borne in mind that, at least in some fields, negative side effects of such measures are to be feared. The need for data transparency is potentially accompanied by the possibility of providing access to production and product information. Once this information becomes public, it is no longer reversible. A unilateral compulsion to disclose provides only very limited opportunities to protect the legitimate interests of the data collector. In certain cases, especially when data collection involves investment and costs, such legislation significantly dilutes incentives to invest. However, this argument does not apply in those cases where there is a business or legal need to collect the data anyway. In conclusion, implementing such a regulation raises the question of what state of data the disclosure requirement concerns: the original "raw data", a structured form (this is how EU 2018/858 is to be interpreted), or information enriched from multiple sources? Such a regime will then potentially affect not only the incentives to collect data, but also the investment in more complex business models based on it. Overall, more insights into the impact of such a law – for instance, based on a careful evaluation of EU 2018/858 – would be desirable before it is implemented within a broad impact framework.

(2) Data Authority, Data Trusts or Platform Solutions

A law with disclosure requirements offers relatively little scope for solutions that consider the different interests of all stakeholders on a case-by-case basis. In a recent analysis of the competition problems in data-based markets in the UK, the appointed expert commission came to the conclusion that government intervention in the provision of data was necessary (Furman 2019). Instead of a legal solution, however, this Commission proposes establishing a specialized authority whose core task would be to resolve conflicts of interest in data access. Such an authority could operate its own data centers and require companies to share their data with these centers. Actors with a legitimate interest in accessing the data could then approach the data authority and make a request to access the data. According to pre-defined criteria, the authority could then decide on a case-by-case basis whether to grant the data access under consideration of the interests of all stakeholders and the resulting welfare effects. This represents a considerable advantage over the legal solution. It must be noted, however, that there is little or no experience on the part of the state in operating data centers with real-time access to massive volumes of data. Moreover, with a large, centralized state data silo, the consequences of a security gap or data loss would be extremely problematic. On the way to implementation, a series of pilot tests would therefore be indispensable, building on the expertise and experience of existing state data authorities (Federal Statistical Office and state offices).

"Data trusts", i.e., private-sector data trustees, represent a private-sector alternative to such a data authority. Similar to the authority, such private companies with the mandate to manage the entrusted data in accordance with a defined charter can make case-based decisions regarding access to the managed information (Mills 2019). In the international context, the UK already has initial experience with pilot projects (ODI 2019). Unlike a public authority, companies must voluntarily submit their data to a data custodian. Accordingly, there is a coordination problem: why should companies share their data, that they would otherwise keep secret, with the trustee? The rationale is that data trustees are a way to solve the prisoner's dilemma in the context of data silos. If two companies operate an own data silo with complementary data, then considerable value added could be created if each had access to the other party's data. Each individual actor, however, has an incentive to keep its own data secret, so that a market equilibrium is created in which no access is granted. If, however, there is an instrument by which both companies can commit themselves to grant each other access, then they are able to break out of the prisoner's dilemma and achieve the allocation that is better for both, in which the data is shared.

This consideration also makes it immediately clear, however, that (voluntary) data trustees cannot solve all the problems associated with data sharing. Firms will generally have no incentive to grant data access to actors via the trustee without any potential economic advantage over the initial situation. Approaches to circumvent this problem are to combine a data trustee with a (commercial) industry platform. Within this framework, access to the data can be granted either in the course of providing the data oneself or through financial participation in the platform. Practical examples show that such approaches are more likely to work if the market players are relatively symmetrical, for example in medium-sized mechanical engineering or across industries in connection with the verification of personal data, and if there is no single dominant player.

In the context of data platforms, the state can play an important role in the design and start-up financing. State involvement also ensures with a higher probability that a critical mass of players can be attracted to the platform. Such an initiative at European level exists in the Gaia-X platform initiative, which is supported by the German government (Handelsblatt 2019).

CONCLUSION

The importance of data economics for developing economic performance in Germany and Europe is undisputed. In this study, we have identified various factors that can cause market failure of data. It is therefore doubtful that the market alone can lead to an efficient allocation of data, and to optimal access to and sufficient participation in the collected data. Initial initiatives, such as the 1996 EU Database Directive, have not had the desired effect on the market. At present, work is urgently underway on possible solutions, whereby state intervention, such as a right to data access, is being pursued in parallel with private sector solutions, such as data trusts or platforms. Due to the complexity of the problem – depending on the extent of external effects, the necessary investment costs and the existing market structure (dominant players or symmetrical market participants) can vary greatly – competition for solutions appears to be desirable. Pilot projects in industries and further research in this area will surely contribute to a better understanding.

REFERENCES

Acquisti, A., C. Taylor and L. Wagman (2016), "The Economics of Privacy", *Journal of Economic Literature* 54, 442–492.

Arrieta-Ibarra, I., L. Goff, D. Jiménez-Hernández, J. Lanier and E. Weyl (2018), "Should We Treat Data as Labor? Moving beyond 'Free'", *AEA Papers and Proceedings* 108, 38–42.

BMWI (2019), *Wertschätzung – Stärkung – Entlastung. Eckpunkte der Mittelstandsstrategie,* August, Berlin.

Brynjolfsson, E., L. Hitt and H. Kim (2011), "Strength in Numbers: How Does Data-driven Decisionmaking Affect Firm Performance?", *Working Paper*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1819486.

Calzolari, G., L. Felli, J. Koenen, G. Spagnolo and K. Stahl (2017), „Relational Contracts, Competition and Innovation: Theory and Evidence from German Car Manufacturers", *Working Paper,* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3035261.

Carrière-Swallow, Y. and V. Haksar (2019), *The Economics and Implications of Data – An Integrated Perspective,* International Monetary Fund. Strategy, Policy and Review Department (Series), September.

Coase, R. (1960), "The Problem of Social Cost", in C. Gopalakrishnan (ed.), *Classic Papers in Natural Resource Economics,* Palgrave Macmillan, London, 87–137.

Deloitte (2019), *Assessing the Value of TfL's Open Data and Digital Partnerships*, http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf.

Duch-Brown, N., B. Martens and F. Mueller-Langer (2017), "The Economics of Ownership, Access and Trade in Digital Data", *Digital Economy Working Paper* 2016-10, JRC Technical Reports.

Economist (2019), "The Internet of Things Will Bring the Internet's Business Model into the Rest of the World", September, https://www.economist.com/technology-quarterly/2019/09/12/the-internet-of-things-will-bring-the-internets-business-model-intothe-rest-of-the-world.

European Commission (2017), *Building a European Data Economy*, SWD (2017) 2 final, January 2018, Brussels.

European Commission (2018), *Towards a Common European Data Space*, April, Brussels, https://ec.europa.eu/digital-single-market/en/news/communication-towards-common-european-data-space.

Falck, O., T. Lohse and J. Alipour (2019), "Transformation zum Dienstleister und neue Anforderungsprofile in der Industrie", *ifo Schnelldienst* 72(1), 43-45.

Farboodi, M., R. Mihet, T. Philippon and L. Veldkamp (2019), "Big Data and Firm Dynamics", AEA Papers and Proceedings 109, 38–42.

Furman, J., D. Coyle, A. Fletcher, D. McAuley and P. Marsden (2019), *Unlocking Digital Competition,* Report of the Digital Competition Expert Panel, March, London, https://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel.

Goldfarb, A. and D. Trefler (2019), "Artificial Intelligence and International Trade", in A. Agrawal, J. Gans and A. Goldfarb (eds.), *The Economics of Artificial Intelligence: An Agenda,* NBER, 463–492.

Handelsblatt (2018), "Tech Startup Celonis Is a Rare Sight in Germany – a Unicorn", 27 June, https://www.handelsblatt.com/today/companies/market-leader-techstartup-celonis-is-a-rare-sight-in-germany-a-unicorn/23582550.html.

Handelsblatt (2019), "Europa-Cloud Gaia-X startet im Oktober", 3 September, https://www.handelsblatt.com/politik/deutschland/datenplattform-europacloud-gaia-x-startet-im-oktober/24974718.html?ticket=ST-12949895-dC2Kf2RZrUwzVa0aJN0lap6.

IDC (2019), *The European Data Market Monitoring Tool Report,* Update of the European Data Market Study SMART 2016/0063, June.

Jaffe, A. (1986), "Technological Opportunity and Spillovers of R&D: Evidence from firms' Patents, Profits, and Market Values", *American Economic Review* 76, 984–1001.

Jones, C. I. and C. Tonetti (2019), "Nonrivalry and the Economics of Data", *NBER Discussion Paper* 26260.

Kerber, W., and J. Frank (2017), "Data Governance Regimes in the Digital Economy: The Example of Connected Cars", *Working Paper,* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3064794.

Koenen, J., S. Krüger and F. Pavel (2018), *Study on Contractual Relationships between Online Platforms and Their Professional Users,* Study for the European Commission (DG Connect), https://publications.europa.eu/en/publication-detail/-/publication/b3d856d9-4885-11e8-be1d-01aa75ed71a1/language-en.

Koutroumpis, P., A. Leiponen and L. Thomas (2017), "Markets for (Big) Data – The (Unfulfilled) Potential of Data Marketplaces", *ETLA Working Papers* 53.

Li, W., M. Nirei and K. Yamana (2019), "Value of Data: There's No Such Thing as a Free Lunch in the Digital Economy", *RIETI Discussion Paper* 19-E-022.

London Economics (2019), *Independent Assessment of the Open Data Institute's Work on Data Trusts* and on the Concept of Data Trusts, April, https://theodi.org/wp-content/uploads/2019/04/Datatrusts-economic-function.pdf.

Mills, S. (2019), "Who Owns the Future? Data Trusts, Data Commons, and the Future of Data Ownership", *Working Paper*, https://ssrn.com/abstract=3437936.

ODI (2019), *Data Trusts: Lessons from Three Pilots*, Report of Open Data Institute, April, https://theodi.org/article/odi-data-trusts-report/.

OECD (2019), *Going Digital: Shaping Policies, Improving Lives,* https://doi.org/10.1787/9789264312012-en.

Schlosser, A. (2018), *You May Have Heard Data Is the New Oil. It's Not,* World Economic Forum Annual Meeting, https://www.weforum.org/agenda/2018/01/data-is-notthe-new-oil/.

Schneider, D. (2011), "Trading at the Speed of Light", *IEEE Spectrum* 48(10), 11–12.

Varian, H. R. (2014), "Beyond Big Data", *Business Economics* 49, 27–31.

Wolfert, S., C, Verdouw and M. Bogaardt (2017), "Big Data in Smart Farming – A Review", *Agricultural Systems* 153, 69–80.