Wirtschaftliche Untersuchungen,
Berichte und Sachverhalte

IW-Report 48/18

On (Intellectual) Property and other Legal Frameworks in the Digital Economy

An Economic Analysis of the Law

Christian Rusche, Marc Scheufen

Köln, 17.12.2018

INSTITUT DER DEUTSCHEN WIRTSCHAFT

Table of Contents

JEL-Classification:

D23 – Organisational Behaviour, Transaction Costs, Property Rights
K11 – Property Law
K12 – Contract Law

Abstract

The significance of data as an economic good in the digital economy quickly raises the question of who owns the data. More specifically, within the context of a data marketplace in which data is being exchanged or traded and where different bundles of rights (property rights) are transferred (contract law) – an answer regarding the genuine assignment of property rights for data seems important.
From an economic standpoint, this paper investigates the need for an (new) intellectual property right for data. Firstly, an overview of the status quo of the literature on property rights for data will be provided. This will be followed by an analysis of the characteristics of data as a good, clarifying the circumstances under which a market failure can occur. Accordingly, only a market failure situation will authorise economic policy intervention. Lastly, we will derive specific policy implications, offering economic reasoning on a new intellectual property right for data and specifically pointing to other forms of law that may be more applicable for reaching an efficient allocation of data resources.

1 Introduction

Data is being called the new oil, but this is an understatement. While consumption of oil is rivalrous, consumption of data is not (Haucap, 2015). The same set of data can be used for various purposes by different people simultaneously. Besides this favourable economic characteristic, digitalisation also eases the collection, storage, processing, distribution and analysis of data, which considerably enhances the role of data for the economy and the society as a whole, by widening the fields in which data can be used. The significance of data is underlined by the fact that companies with data-driven business models already dominate the ranking of the most valuable companies worldwide (Rusche, 2018 Forthcoming). Furthermore, the importance of data is still increasing. In a study for the European Commission, it is estimated that the data market in the European Union (EU) could have a total impact on GDP of around EUR 739 million in 2020, which is an increase of more than 25 percent since 2016 (IDC/Open Evidence, 2017).

In light of the considerable role played by data, discussion arose about how to provide the economy with the data needed to enhance growth and enable new data-driven business models. In this context, the matter of mandatory data sharing is also a topic of discussion (Haucap et al., 2018, 172). However, there tends to be more than one agent involved when data is created. For example, a company producing data using machines with sensors built by another company, where the latter could, during maintenance works, also access the data, or consumers using a search engine on their mobile device, while the various companies who provided the hardware or software required to make the search may have access to the search query data, too. In this respect, the questions of whether or not there is an "owner" of the data at hand or whether or not there should be a new intellectual property right for data surfaced. These questions were also raised by the European Commission. In the Digital Single Market Strategy, the Commission mentioned that without clarity over rights to use data, the single market could be fragmented and new technologies could be hampered due to differing regulations in the Member States (European Commission, 2015, 14).

Regarding the clarity over rights to use data, authors like Drexl (2017) and Franceschi/Lehmann (2015), among others, argue that the existing laws regarding intellectual property can be used to address access to, and ownership of, data. In contrast, Zech (2016) is in favour of a new intellectual property right for data. Generally, an intellectual property right is a cluster of capacities that varies according to the intellectual property at hand (Ohly, 2014, 3). In other words, an intellectual property right (IPR) for data is only needed if it solves specific (economic, contractual or other) challenges that arise due to data. This article contributes to the discussion by analysing the characteristics of data and the market failures that may be caused by these characteristics. We will then address the data-specific problems that the new IPR (as well as other forms of law) would have to solve.

With respect to the question of ownership, a distinction must be made between personal and non-personal data. Personal data refers to "any information relating to an identified or identifiable natural person" (Article 4 (1) General Data Protection Regulation (GDPR), Official Journal of the European Union, 2016a). According to Zech (2016, 464), the data protection regulation

does not aim to provide a property-like regulation for personal data. In fact, it tries to introduce a personality right that balances the different interests. Since the European Union created a level playing field regarding access to data and the purposes that they can be used for in the internal market by implementing the GDPR, this paper only concentrates on non-personal data, especially machine data. This non-personal data also plays a crucial role for the European economy. For example, data allows for the creation of a digital twin of a process in the analogue world, which can then be analysed, changed and optimised without affecting production (Lichtblau et al., 2018).

Nevertheless, the scope of data that is regarded as personal data is broad. The Court of Justice of the European Union (2016b) even defined a dynamic IP address as personal data. The reason being that the individual in question can be identified through a combination with the data of the provider. However, thanks to proper pseudonymisation, as mentioned in the GDPR, or by using synthetic data based on the personal data at hand (Drechsler/Jentzsch, 2018), a data set can be transformed to non-personal data.

This paper proceeds as follows. In section 2, an overview is provided of the current legal literature on ownership of non-personal data. This includes the regulations that can be applied as well as the discussion of whether or not there should be a new IPR for data. In section 3, a summary is given of the economic characteristics of data and the possible causes for market failure. On that basis, policy implications are discussed in section 4, before the conclusions drawn in section 5.

2 Literature Review on Ownership of Data and Access to Data

Because of the high and ever-increasing importance of data, the ownership of, and the right to access, data have been discussed in particular in legal and economic literature. Evidence that this discussion is not limited to a mere theoretical analysis was provided by the Digital Single Market Strategy of the European Commission, which outlines possible legal actions – regarding, among other things, data ownership and a right to access data – aimed at helping establish a Digital Single Market within the European Union (Zech, 2016; European Commission, 2015, 15). As yet, however, such rights have not been introduced.

On the one hand, this gave rise to different interpretations of other intellectual property rights and further regulations with respect to data. In section 2.1.1, interpretations that go back to EU regulations and national regulation in Germany are discussed. On the other hand, a discussion evolved that tries to evaluate whether a new intellectual property right (IPR) for data is needed or not. And, if it is needed, what it should look like. We will discuss the literature on this topic in section 2.2.

In this context, however, it has to be considered that data are used to extract the valuable information they contain (Rusche, 2018 Forthcoming). Accordingly, data can be regarded as a raw material that has to be utilised. This connection can also be found in the legal discussion. In fact,

there is a distinction made between the syntactic and the semantic level (among others by Drexl, 2017 and Zech, 2016). The syntactic level refers to the characters alone, e.g. the letters of a text. The semantic level refers to the meaning, e.g. the statement of a text. In this article, we will also apply this distinction and concentrate on data on the syntactic level. The reason is that an additional investment is necessary to extract valuable information (semantic level) from the sequence of "zeros" and "ones" (syntactic level). This holds especially true for data that is generated automatically by machines.

2.1 Property Rights Related to Data

The right to access data or data ownership can be subject to the contracts concluded by the parties involved, such as companies (Ensthaler, 2016, 3474). However, it has to be clarified which legislations may apply if such a contract is missing. Furthermore, a defined ownership can be a starting point for negotiations and therefore lays the ground for a contract. Once the ownership has been clarified, it is also possible to regulate access to data, for example by introducing compulsory licenses (Weber, 2017).

2.1.1 Union Wide

As mentioned above, it is the aim of the European Commission to foster a single market for data. To this end, legislations are introduced to unify the laws of the Member States. In this section, we will deal with the European legislations that may apply to data, and therefore be applicable across the European Union. If EU legislation is missing, however, the national legislation of the Member State may apply. As an example, in section 2.1.2. we will discuss the approaches that relate to German law.

Database Directive

One piece of legislation that is applicable to data is the law on databases. This legislation goes back to European Directive 96/9/EC (Database Directive) published in 1996 (Official Journal of the European Communities, 1996).[1] Within the framework of this Directive, a database is "a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means." (Art. 1 (2.)).

This definition already shows that a database is not only a mere collection of data (syntactic level). In order to create a database, the data and information have to be arranged in a systematic way (semantic level). Accordingly, it is not the aim of the regulation to protect the single data points within the database. The aim of the regulation is to protect the investment made to establish the database (Duisberg, 2017, 24). This also includes the investment in human capital for finding a suitable structure for the database. However, investments for producing data (ibid.) are not included.

[1] The Directive was implemented in Germany in §§87a et seqq. Urhebergesetz (copyright act)

In essence, the Database Directive only provides for an exclusive property right for the whole database. A single data point within the database is not protected. The most important reason behind this is to not exclude the possibility for competitors to produce, or collect even, the same data on their own and create their own database.

Trade Secrets

In 2016, the European Union adopted Directive (EU) 2016/943 to harmonise the regulation of trade secrets within the Single Market (Official Journal of the European Union, 2016b).

According to the Directive, a trade secret has to fulfil all of the following three requirements (Art. 2 (1.) Directive 2016/943):

"(a) it is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question;
(b) it has commercial value because it is secret;
(c) it has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret".

It should be noted that this regulation considers the acquisition of a trade secret as lawful if the discovery or creation took place independently (Art. 3 (1) lit. (a)). Accordingly, the holder of a valuable set of data has no right to exclude any other person or company from legally creating and using a data set that is as valuable or even identical. Thus, the trade secret regulation does not provide exclusive rights on the trade secret (Kiefer, 2018, 912). Furthermore, the fact that a data set is not publicly available and that the holder has made investments to keep it secret lays the ground for considering a data set as a trade secret according to Directive (EU) 2016/943. That also means that once the data is disclosed, it loses the status as a trade secret, and therefore loses value. However, the Directive grants protection to the holder of the data in the event of unlawful acquisition, use or disclosure thereof (Art. 12). This also includes the payment of damages suffered by the trade secret holder (Art. 14).

The compensation for the damages suffered can consist of three parts: direct monetary loss, the profit earned by the company that unlawfully acquired the information and a (virtual) license fee for the trade secret (Kiefer, 2018, 914).

Antitrust regulation

In this respect, competition problems due to a lack of access to data or market power because of data can be addressed by antitrust regulation as well. This applies for merger control (Rusche, 2017) and access to data of a dominant company (Rusche, 2018 Forthcoming). However, as is stated by Hennemann and Paal (2018, 81), the allocation of ownership of data and access rights to data are necessary preliminary steps for determining the market position, and hence market power. Accordingly, consequences of antitrust regulation are discussed in section 4.2.

2.1.2 Germany

Due to a lack of a regulation that unifies ownership of, and access to, data in the European Union, a great deal of discussion evolved that tried to apply existing regulations to data. In this section, we describe the approaches taken in Germany.

One string of literature attempted to apply laws from the German Civil Code (Bürgerliches Gesetzbuch - BGB) that deal with physical goods.[2] This included an approach that uses §§903 S.1, 90 BGB (Zech, 2015). This approach defines the producer of data as the holder of property rights. Producing data in this respect means recording or coding (Duisberg, 2017, 19). Accordingly, by collecting data, a party can claim rights to the collection.

Hoeren (2013) describes a rather similar approach. He refers to § 303a StGB from the German Penal Code (StGB – Strafgesetzbuch). According to his interpretation of the law, the person that technically created the data, either by directly saving it in a data memory or using a programme (Hoeren, 2013, 487), can claim ownership. It should be noted that the owner of the data memory is not the owner of the data, but he may have a right to access the data.

Hoeren (2013) also attempts to draw similarities between data and physical goods, and the respective relevant laws. In this respect, he also refers to §950 BGB. This paragraph is also used in another approach by Ensthaler (2016). According to this approach, the right of ownership of data could be decided analogously to the handling of raw materials (Ensthaler, 2016, p. 3475). Ensthaler (2016) describes the example of a company that uses machines: the manufacturer installs devices that collect data from these machines. The company using the machines produces the raw data and can therefore charge a (small) fee, but the producer that installed the devices can claim the rights to the information generated from the data and the databases created from these data. Furthermore, according to Ensthaler, the fee is supposed to be unrelated to the value that can be created by utilising the data.

Contrary to this approach, Grosskopf (2012) argues that machine data is a product of the machine the data is generated from and therefore belongs to the owner of the machine ("Früchte einer Sache", §99 BGB). However, Grosskopf (2012) limits his analysis to geodata and telemetry data. Furthermore, rights may only come to live if the yield of data is as intended, which may vary according to the case at hand.

Another approach neglects proprietary rights for data (Assion/Heun, 2015; Duisberg, 2017). However, Assion/Heun (2015) use §100 BGB to argue that the value created by utilising the data can be allocated to the owner of the device on which the data is saved. In other words, the value created stems from use of the device, and thus the device owner has the right to the value created. However, the authors state that ownership in this sense may be disputed, for example if data is saved in a cloud.

[2] We will elaborate in section 3 that the characteristics of physical goods do not apply to data from an economic perspective.

2.2 Intellectual Property Rights for Data

Against the backdrop of the abovementioned approaches, the question arises as to whether there is still a need for a new intellectual property right with respect to (syntactic) data or not. In fact, do problems of market failure exist that warrant regulatory action? In the following subsection, a summary is provided of the discussion of this matter in the relevant literature.

2.2.1 Discussion of the Need for a New IPR

Franceschi/Lehmann (2015) conclude that the existing rules can already be used to regulate the data market. However, they only describe regulations in Germany and Italy. Kerber (2016) denies the need for a new intellectual property right, although he does state that the development of the data economy could make the introduction of additional laws necessary soon. Kerber's denial of the need for a new intellectual property regulation is firstly based on the notion that the cooperating parties at hand know their business best, and can thus agree on suitable contracts that address the allocation of value added and the access to data. The problems that arise, for example due to market power of one party, can be addressed by existing laws, i.e. competition law in this case. Secondly, Kerber is not convinced that too little data is produced and analysed because there are suitable means for keeping data secret and to prevent unauthorised copying. Accordingly, there is no public good problem and the development of a market for data is not hampered.

Finally, Kerber believes that a new intellectual property right causes damages because there are already "difficulties of defining and specifying the exact subject matter" (Kerber, 2016, 998). Furthermore, the development of the data economy cannot be foreseen in detail and therefore an inappropriate regulation could impede economic development and cause legal uncertainty. This argument is also used by stakeholders in the economy, such as the Federation of German Industry (BDI, Bundesverband der Deutschen Industrie, 2015) and the Plattform Industrie 4.0 (2016), an initiative that is supported by politics, business, science, associations and unions to promote Industrie 4.0, especially in Germany. In fact, both are in favour of contracts concluded by business partners and against a preventive introduction of a new intellectual property law.

Drexl (2017) also believes that the existing rules are sufficient and that there is in principle no need for a new regulation. However, Drexl (2017, 291) sees uncertainties in applying existing rules to data, such as the Trade Secrets Directive or EU competition law, and the clarification of these uncertainties by courts might not take place quickly.

However, there are also arguments in favour of a new intellectual property right. Zech (2016) mentions that a uniform right may help to establish an EU Single Market. Furthermore, Zech (2016, 470) argues that rights help to establish a functioning market for data. Property rights make companies offer data sets on markets that they would otherwise have kept secret. Another stakeholder in favour of a new regulation is the German Federal Ministry of Transport and Digital Infrastructure (Bundesministerium für Verkehr und digitale Infrastruktur, 2017). The Ministry aims to introduce a new regulation that clarifies access rights to data. This would help

to make more data available and therefore to utilise the opportunities data offers. However, it must be said that only a minority of authors are in favour of a new IPR (Lohsee et al., 2017, 18).

2.2.2 Characteristics of a New IPR

As mentioned previously, an intellectual property right is usually a cluster of capacities that varies according to the intellectual property at hand (Ohly, 2014, 3). A trade secret, for example, incorporates capacities other than trademark rights. While trade secret regulation aims to protect the effort put into keeping a secret secret, trademark rights ensure that the holder of the rights can exploit the public trademark he created.

When considering intellectual property rights, there are four important features that must be factored in and that may differ according to the intellectual property in question (Kiefer, 2018, 911). They are subjectivity, totality, exploitability and the scope of protection, and they will be discussed with respect to data further on. Prior to that, however, it must be mentioned that defining the subject matter of an intellectual property right for syntactic data is problematic in itself (Kerber, 2016, 998). For example, Drexl (2017, 277) argues that such a right is questionable because it has to be shown that the data is new. In other words, it must be demonstrated that "the same information has not been stored before in form of 0s and 1s." (Drexl, 2017, 277, fn 109). Furthermore, one must consider that the same information can be stated differently on the syntactic level, in a different language or form, for example. Hugenholtz (2017, 97) mentions that defining the subject matter for real-time data might even be impossible. In fact, the question generally arises of what the consequences would be if a new data point were added to a data set that was protected by an intellectual property right: is that already sufficient grounds for creating a new intellectual property right?

- Subjectivity

  Once the subject matter has been defined, however, an intellectual property right for this subject matter has to include subjectivity. In fact, intellectual property rights must be assigned to a subject (Kiefer, 2018, 911). This subject is either a natural person or a legal entity. The subject is the holder of the intellectual property right. The holder of an intellectual property right is granted the opportunity of exclusively using the intellectual property he created or acquired. The question with respect to data is, to whom should the right be allocated? As explained in section 2.1, different approaches allocate the right to different subjects. Zech (2016), for example, is in favour of attributing rights to the data producer, while Assion/Heun (2015) allocates the right to the owner of the device on which the data is saved.

- Totality

  According to Jänich (2002, 198), totality refers to the fact that once the intellectual property right and the right holder are defined, the general public has to respect this decision and is not allowed to impede the holder of the right. According to Kiefer (2018, 913), totality is an exclusivity in relation to third parties that is expressed by the right to use the intellectual property and the right to defend oneself against unjustified interventions. Since the degree

of totality is an essential feature of an intellectual property right for data, we will hold back on analysing this point until section 4.1.

- Exploitability

  Exploitability is the right to make use of the intellectual property. In other words, the holder of the intellectual property right can use it for his own business, sell it, or give licenses to a licensee and therefore grant a contracting party the right to monetarise it. That this feature also applies and should apply for data is not doubted in the literature. Furthermore, compulsory licenses are mentioned in the literature to address the problem of missing access to data (Weber, 2017).

- Scope of Protection

  This feature addresses not only the scope of protection of the right itself, but also the possibilities the holder of an intellectual property right has to defend himself against a violation of his right. The scope of protection summarises aspects such as the limitation period of the exclusive right, which actions are prohibited for competitors or to whom the holder should grant access to the right. Defensive rights could include indemnity for a violation of the right or injunctive relief. However, since the majority of authors argue against the introduction of a new intellectual property right for data (section 2.2.1), specific configurations are not addressed. Zech (2015, 2016), who is in favour of a new intellectual property right, has addressed the question of the limitation period and concluded that there should only be a "short-term protection" (Zech, 2016). The rights of defence with respect to data may be in line with the rights granted in Directive 2016/943 (Official Journal of the European Union, 2016b) on trade secrets, mentioned in section 2.1.1, and that are also in line with the rights granted for other intellectual property rights (Kiefer, 2018). In fact, once the right is violated, the holder is entitled to compensation. Furthermore, in Art. 12 of the Directive (Official Journal of the European Union, 2016b), the holder may call for:

  - the cessation of or the prohibition of the use of or disclosure of the trade secret;
  - "the prohibition of the production, offering, placing on the market or use of infringing goods, or the importation, export or storage of infringing goods for those purposes";
  - the recall, depriving or destruction of infringing goods; the destruction of all or part of any document, object, material, substance or electronic file containing or embodying the trade secret.

3 Intellectual Property Rights for Data: An Economic Approach

Analysing the need for a new (intellectual) property right for data requires an in-depth understanding of the general characteristics of data. In particular, we classify goods as rivalrous/non-rival and excludable/non-excludable. Here, once we have categorised data in such a way, we will discuss possible forms of market failures. Looking at specific conditions that must prevail for an efficient market outcome, we will identify and isolate various market failure issues that may be predominant in a market for data. Finally, we will investigate under which conditions different forms of market failure may be most likely, making a distinction between different types of data.

3.1 Characteristics of Data

In general, two characteristics can be distinguished to classify goods from an economic perspective. First, goods can be described in terms of rivalrous versus non-rival in consumption. A good is (non-) rivalrous in consumption if the use of the good by one person does (not) impede the use of the same good by another person. For example, drinking a glass of water implies that nobody else can drink the very same glass of water (i.e. water is rivalrous in consumption). In contrast, using the German autobahn does not impede other drivers from deriving the same utility from its usage (i.e. the autobahn is non-rival in consumption). The latter example, however, reveals that a good can change its characteristic from non-rival to rivalrous in consumption and vice versa. Accordingly, a traffic jam would induce rivalry in the use of the autobahn.

Secondly, goods can be characterised by excludability or non-excludability. Excludability implies that others can be excluded from using the good, whereas non-excludability means they cannot. For instance, owning a car and having its key enables its owner to exclude others from driving the car (i.e. a car is excludable in consumption). In contrast, it is not possible or extremely costly to prohibit others from fishing in the ocean (i.e. fish are non-excludable in consumption). Classifying goods in terms of both rivalry and excludability in consumption reveals a matrix containing four different types of economic goods (Table 3-1).

Table 3-1: Classification Characteristics of Economic Goods

	Excludability	**Non-Excludability**
Rivalry	private good	common good
Non-Rivalry	club good	public good

Source: own depiction

Accordingly, having (exclusive) ownership of a certain good necessarily defines whether others can be excluded from its consumption. That is to say, property defines a private or club good, depending on whether the good is rivalrous or non-rival in consumption. In contrast, for every good where no property right is defined, the non-excludability characteristic reveals a common-

pool resource (given rivalry in consumption) or a public good (given non-rivalry in consumption). Thus, the need for a (possible) property right on data requires clarification of the genuine characteristics of data as a good.

Firstly, data is characterised by non-rivalry in consumption. As data cannot be depleted, it can be consumed by more than one party without any loss of individual utility through consumption. From an economic perspective, this characteristic is beneficial, as it reveals that data is not a scarce resource. Hence, existing data should be used by as many consumers as possible, as from a social welfare perspective, each net benefit from data consumption increases social welfare. As an interim conclusion, we find that data is either a club or a public good, depending on the question as to whether (exclusive) property on data can be defined.

Secondly, data can be characterised by excludability, as there are many ways for a company's internal data governance to prevent data from becoming public. For example, a company can determine roles for employees, such as data owner or data steward. For such internal use, data governance can be structured according to two main forms of organisation. On the one hand, data governance may be characterised by a centralised ownership model with one owner of all data resources. On the other, a decentralised ownership model may specifically define personal roles of different owners of explicit and distinct data resources. Accordingly, for internal use, contractual employment provision combined with technical systems (such as authentication- and authorisation-protocols) may allow for excluding third parties from using a company's data.

In contrast, the increasing importance of externally used data – especially as a good that can be traded on a market for data – may challenge the excludability characteristic of data. The reason for this can be illustrated by the well-known information paradox by Kenneth Arrow (Arrow, 1962). For experience goods in particular, the seller of a good has to provide information about the product for the buyer's (pricing) assessment. However, once this information is disclosed, the buyer cannot be excluded from this information. This is particularly a problem for information goods, such as data. In other words, it may be questionable whether or not third parties can be excluded from using external (traded) data, revealing both characteristics of a club as well as a public good. In this regard, Kerber (2016) argues that the information paradox may not be a substantial problem for data, as it can be assumed that for most cases a seller of data will be able to sufficiently describe the value of his data to a buyer without disclosing the full set of data. However, Kerber (2016) points to another closely related problem: the buyer of data may be able to resell the data to others with potentially very negative effects on the expected revenue for the initial seller (section 3.2.3). For data with high production costs in particular, this aspect may be crucial, as the buyer benefits from substantial cost advantages by copying the data and reselling it to other potential buyers.

3.2 Market Failure Arguments

Whether or not we conclude an efficient allocation of (data) resources is always a question of identifying a possible market failure. Economic theory tells us that an intervention (e.g. by the definition of a(n) (intellectual) property right for data) may only be justified by the existence of a market failure. To identify a market failure, neoclassical theory typically refers to the model of

perfect competition as a benchmark for (economic) efficiency. Accordingly, a set of requirements has to be met to allow for an efficient allocation of data resources. Table 3-2 provides an overview of these requirements as well as the respective market failure in the event of violation.

Table 3-2: Perfect Competition Assumptions and Market Failure

No	Assumption	Market Failure
1	Well-defined property rights	Externalities
2	Absence of transaction costs	
3	Absence of externalities	
4	Atomistic market Structure	Market power
5	No market barriers (entry/exit)	
6	Absence of economies of scale and scope	Natural monopoly
7	Perfect information	Information asymmetry (i.e. adverse selection and moral hazard)
8	Homogenous products	

Source: Extending on Fritsch (2014, 26) and Scheufen (2018, 220)

Obviously, the characteristics of, and the market for, data may question some of the assumptions of perfect competition, necessarily inducing a market failure debate and manifold options for market intervention. Before we look more closely at the question of proper policy intervention, we will discuss possible market failure problems in the context of data, starting with externalities (and hence a violation of assumptions 1 to 3), followed by problems of market power (assumptions 4 and 5) and natural monopoly (assumption 6), and lastly information asymmetry in the context of data due to a violation of assumptions 7 and 8.

3.2.1 Externalities

Obviously, questioning whether or not data can be characterised by excludability or non-excludability crucially links to assumption number 1, i.e. well-defined property rights. From economic theory, we know that well-defined property rights are particularly important for an efficient incentive setting in the context of a commons world. The term commons refers to any good that can be characterised by non-excludability, i.e. public and common goods. In this regard, we can learn from the commons and anti-commons literature (e.g. Heller, 1998; Buchanan/Yong, 2000). On the one hand, a common good is typically characterised by non-excludability and rivalry in consumption (Table 3-1). The non-excludability characteristic triggers an incentives setting where the initiator of an action does not consider the negative externality caused to a third party due to rivalry in consumption. For instance, a hunter shooting an animal gains from his hunt, while not considering the negative effect caused for the population and hence the chance of his friend to get a good hunting result. If there is no well-defined property right for (hunting) animals, hunters would have the incentive to shoot as many animals as possible, leading to overhunting,[3] i.e. the incentives problem is characterised by overprovision. On the other hand,

[3] Please note that property rights do not necessarily have to be defined externally, but can also emerge spontaneously. A famous example is Demsetz' (1967) study on the native inhabitants of the Labrador Peninsula. Here, the

non-excludability and non-rivalry in consumption leads to the well-known public good problem, i.e. free riding on the positive externality of a non-excludable good and hence an incentives problem defined by an underprovision. In the context of data, the latter incentives problem may apply.

The economic literature (e.g. Kerber, 2016; Duch-Brown et al., 2017) mostly discusses the incentives problem for data based on two arguments. First and foremost, other means of excludability (especially contractual solutions combined with technical restrictions) may prevent an underprovision of data. Secondly, the empirical fact that a great amount of data can be produced at very low (or even zero) cost (Rifkin, July 2015) will likely prevent a negative cost benefit analysis on the creator level and hence an underprovision. Despite the empirical evidence that this characteristic may apply for the majority of (in part freely available (open)) data, the latter argument may depend decisively on sectors and may particularly differ for a variety of emerging business models that are defined to some extent by decisive investments in the collection and processing of data.

Another consideration closely related to an incentives problem is a possible violation of assumption number 2, i.e. the absence of transaction costs, which may prevent a Coasian[4] solution for the externalities problem in the market for data. In this regard, especially for machine-generated (e.g. sensor) data, it may often be unclear who the owner of the data is. For example, a company integrating sensors into machines for offering additional services may question whether the original creator of the machine is also the producer of the machine-generated data. There are in fact several examples for which multiple creators or holders of data exist (e.g. connected cars, where the driver, the manufacturer or suppliers can create and/or consume data), leading to possibly prohibitive transaction costs and hence externality problems, e.g. due to a large number of contracting parties.

Finally, Kerber (2016, 993) also points to the fact that there can be positive and negative externalities (assumption number 3) in the production, analysis and use of data. From an economic perspective, the existence of such externalities would create a gap between private and social benefits and costs, and hence result in an inefficient allocation of data resources. On the one hand, the presence of positive externalities would lead to a situation similar to the public good or incentives problem discussed above. That is, the creator of these positive effects resulting in the production, analysis or use of data would be unable to internalise the social benefits, leading

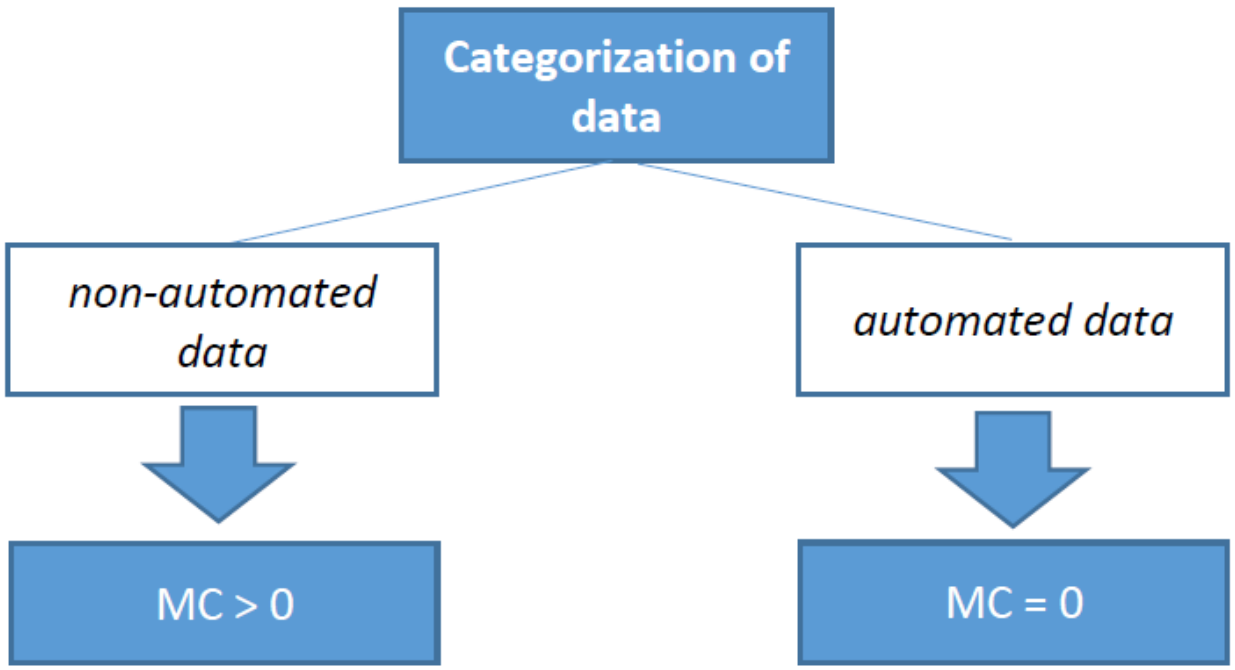
immigration of Europeans in the eighteenth century led to an increase in the demand for fur-bearing animals (especially beavers). Without property rights on beavers, we would have expected to observe overhunting. Demsetz, however, finds a spontaneous emergence of property rights in terms of hunting territories, as beavers are characterised by relatively small habitats and hence a low cost of internalisation. In each hunting territory, the native inhabitant finally has an incentive for a sustainable hunting activity, preventing him from externalising his costs of hunting.

[4] The Coasian solution reverts back to the famous Coase-theorem. Coase (1960) finds that in the absence of transaction costs and given well-defined property rights a bargaining solution between the private parties involved leads to an efficient allocation of resources.

to private incentives that are lower than the socially optimal ones.[5] On the other hand, the presence of negative externalities would lead to a situation similar to the commons problem cited above, i.e. private marginal cost would be smaller than social marginal cost (with the difference determining the extent of the external costs), resulting in an allocation that is above the socially optimal level of data production. In fact, it may be the type or characteristic of the data at hand that determines whether a market failure exists and in which direction (over- or underprovision) the pendulum will swing.

While Kerber (2016) and Duch-Brown et al. (2017) find no evidence for an incentives problem in the market for data and hence no need for a new IPR for data, this may depend greatly on the production costs of data. In fact, Kerber (2016) argues that most data can be produced at virtually no variable cost, likely resulting in an incentive to create new data, since the benefits are likely to exceed its cost, where marginal returns (MR) ≥ marginal cost (MC) = 0. However, this may depend decisively on the kind of data we are looking at. Several new business models in the market for data are in fact characterised by (partly) excessive production costs. To our knowledge, there is no universal categorisation of types of data except personal versus non-personal data. However, taking the creator of the data into account, we are able to distinguish between data collection involving human capital (henceforth non-automated data) and automatic collection of data by machines (henceforth automated data) (Figure 3-1).

Figure 3-1: Types of Data – A Categorisation



Source: own depiction

[5] For instance, the owner of a nice front yard does not take into account the positive impact of his garden on his neighbours (from enjoying the nice view), while a subsidy from his neighbours for the positive externality could be used to make the garden even more beautiful.

Accordingly, Kerber's argument that the cost of producing data is very low or zero may hold particularly true for automated data, i.e. data produced via sensors integrated into machines. Despite a comparatively low investment for the sensor technology and an adequate database[6] to collect the data produced automatically, there are no decisive sunk costs. By collecting data en masse from machines, for example to offer additional services to buyers of production technologies, the marginal cost for a single data point is close to zero.

In contrast, non-automated data may be characterised by higher, and in some instances decisive, investments in data production. For example, a company collecting comprehensive network data on environmental aspects (for example, data on temperature, humidity, solar radiation, traffic light intervals and potential traffic congestion) may face high investments for data blackbox solutions, vehicles and respective costs for drivers and petrol. In fact, the environmental data itself may be available free of charge. However, the data collection process, e.g. for a comprehensive network of environmental data to cover all parts of a city, region, country or continent, may be costly. However, any kind of (non-automated) data that comes accompanied by high investments, establishes a basis to challenge the incentives for the data creator, as the creator seeks to recoup his investment at least, from collecting and trading data. Obviously, it is an empirical question how relevant the two types of data are (market share of such data) and how problematic this incentives problem may be. Moreover, one can still question whether a new IPR is the right way to correct such a market failure, or whether there are other economic policy means for intervening with respect to the market mechanism.

3.2.2 Market Power and Natural Monopoly

As data is seen as the most valuable resource of our century, it does not seem far-fetched that a mono- or duopolistic market structure (assumption number 4), as well as the existence of market barriers (assumption number 5), may lead to market power as another market failure problem in the market for data. In fact, seven out of the ten most valuable companies can be assigned to data-driven companies (Rusche, 2018 Forthcoming, 1), greatly revealing the existence of (at least financial) market power in the data-driven economy. The abuse of this market power (e.g. by impeding access to data as an essential facility) may create market barriers that prevent effective competition. On the other hand, data may facilitate competitive practices (e.g. by enhancing market transparency), but may also promote price discrimination, for instance, by using consumer data to differentiate between different types of consumers or markets. In fact, an IPR for data would also automatically induce market power and hence a market failure, preventing an efficient allocation of data resources.

Last but not least, it may not always be clear whether the actual situation of data holders is already efficient. In many respects, data markets may be characterised by network effects and a subadditive cost structure (i.e. due to economies of scale (assumption number 6 in Table 3-1) an efficient allocation may only be reached if one company produces the data), leaving us with a market failure of market power stemming from a natural monopoly. The reason for this is that, similar to train tracks or other infrastructure, it is not efficient for specific forms of data to re-

[6] As we have seen a database is protected (section 2.1).

create existing data infrastructures, but rather to foster access to such monopolistic bottlenecks[7] to enable competition on upstream and downstream markets. For example, when collecting data on outdoor temperature, it is not efficient for every company to build his own weather station, but instead to use the existing infrastructure of weather stations. As we will argue in section 4.2, such market characteristics may reveal that most market failure problems arise due to unclear regulations on data access.

3.2.3 Information Asymmetry

Besides externalities and tendencies towards market power/natural monopoly, problems of asymmetric information could also be present in the market for data, questioning assumptions such as perfect information (assumption number 7) and homogenous products (assumption number 8). The assumption of homogeneity of different types and sets of data, in particular, can obviously hardly apply here, as each data set is unique in the sense that the database structure and hence analytical applications or tools may vary considerably between different types and sets of data. In this regard, to a certain extent specific data sets may, like academic journals, bundle different information to form a unique set of information (i.e. an issue of a journal). As different journals form different sets of information, a university library seeking to offer complete access to academic knowledge will act as a price taker, i.e. it will basically have to subscribe to all journal sets of their field, since different journals are not substitutable. From a market structure perspective, this forms a setting of monopolistic competition (Eger/Scheufen, 2018, 15 ff.). Similarly, matching different sets of data will likely put data-driven companies in a position to not only ensure interoperability, but for a credible signalling of their data characteristics and quality to also prevent adverse selection. The latter argument necessarily links to the assumption of perfect information. Actually, a credible signalling of different data characteristics and their quality raises the intriguing question of how to assess data quality, as well as the value of data. The literature on the assessment of data is still seeking a solution to this problem (e.g. Heckman et al., 2015). In fact, there is no comprehensive approach so far as to how to assess the quality and the value of data, especially since this crucially depends on the purpose for which the data is intended, which may vary considerably between users. Consequently, a signalling and/or screening approach to overcome the information asymmetry between buyer and seller of data is likely to fail.

Varian (2003, 674) refers to adverse selection as a problem of hidden information. However, with asymmetric information there might also be a problem due to hidden action, which is known as moral hazard. This problem arises in particular in insurance markets and describes a situation where the actions of the consumer affect the probability of causing damages. The consumer's behaviour with insurance protection might be riskier than without, due to the fact that the insurance company pays for the damages and cannot observe the behaviour of the consumer.

[7] Monopolistic bottleneck is a term used in competition law to describe an essential facility that is necessary to operate on a market and cannot (or due to prohibitive costs) be created independently. For instance, offering rail services requires access to train tracks, while creating a train track network independently is not efficient.

Moral hazard, however, can also exist in the data economy. Data are non-rival and can be copied and distributed further at virtually zero cost. Accordingly, there are incentives for companies that have bought a license to access data in order to resell the data, for example. Even if the passing on of data was excluded in the contract, however, a new product could be developed by recombining or matching the data, which could then be sold. This would limit the market and therefore damage the company that granted access to the data in the first place. The fear that this might happen works as an incentive to keep the data secret, inducing a market barrier for other companies seeking to exploit the data by means of a new business model.

The fear of moral hazard can also lead to high investments in technically securing data, or in producing synthetic data. Both would increase the price charged for access by the holding company and therefore lead to the scenario that demand at the higher price would decrease.

4 Policy Implications

Having set the ground for understanding under which conditions the market for data is inefficient, we will derive different policy implications to correct specific forms of market failures. Firstly, we will investigate the need for, as well as the optimal design of, a new intellectual property right for data, benefiting from our insights into the market failure of externalities in section 3. We will challenge various features of such a new IPR, as discussed in the legal literature from an economic point of view. Secondly, we will point to possible linkages to other forms of law, such as competition law (to correct the market failure of market power and/or natural monopolies) or contract law (to correct the market failure of asymmetric information), which may be affected (also as a result of a new IPR) in a market for data.

4.1 Externalities: IPR as a Solution

4.1.1 The Economics of IPR

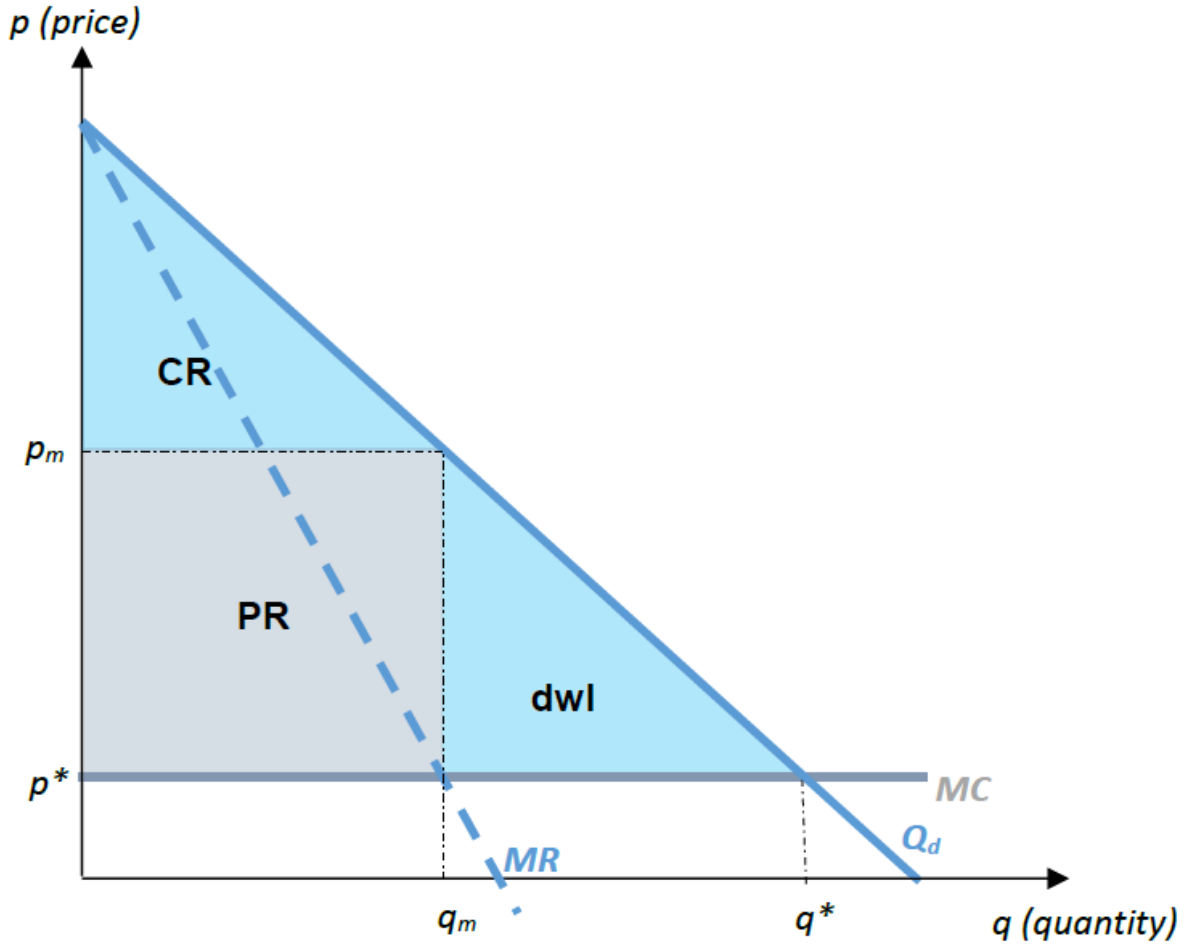
In general, the economics of IPR argue that, due to the public good nature, an incentives problem arises that may only be solved by introducing an exclusive right to protect against copying or imitation. The exclusivity of such a right differs considerably from one form of IPR to another. While a patent grants exclusivity to the idea of an innovation (i.e. very broad understanding of exclusivity) and hence gives priority to the innovator, copyright only protects the expression, allowing for double creation as well. The line of argumentation is twofold: first of all, the public good nature of the information good leads to free riding on the investments of the innovator. As a public good is defined by non-excludability (and non-rivalry), nobody would have an incentive to pay for the good. Despite some first-mover advantages[8], the missing incentive for the

[8] A first-mover advantage results from being the first to be able to produce and sell the product. Possible imitators of the innovation will have to create the absorptive capacity (Cohen/Levinthal, 1989) to enter the market. Depending on the complexity of the product, this absorptive capacity may determine the degree to which the innovator may benefit from being the first player on the market.

innovator would lead to an underprovision of innovation. By introducing an IPR for the innovation good, the policymaker seeks to provide an incentive to invest in research and development (R&D) as the innovator receives an exclusive right to produce and sell his good. Accordingly, the temporary monopoly granted to the innovator, and hence the monopoly rent gained from setting profit-maximizing prices, should provide enough (financial) incentives to invest in R&D (Scheufen, 2015, 21 ff.). Figure 4-1 illustrates the typical model used to visualise the static impact of a patent for a new product innovation.[9]

Figure 4-1: Welfare Effects in an IPR System

Example: Patent



Source: own depiction

Given a marginal cost function (MC) and a demand function (Q_d), a profit maximising innovator holding a patent sets a monopolistic price (p_m), yielding marginal cost equal to marginal revenue.

[9] Please note that innovations do not always generate a new product, but sometimes involve process innovations. Process innovations are typically illustrated by modelling the cost effect of a new process innovation by means of a shift in the marginal cost function. Whether or not the process innovator is able to set monopolistic prices, depends on whether the monopolistic price is above or below marginal costs before the process innovation was introduced. In contrast to a product innovation, it is not a new market that is created (with consumer, producer rent as well as dead-weight-loss), but only an efficiency or cost advantage as depicted by the distance between the marginal cost function before and after the process innovation.

Compared to the efficient market equilibrium under perfect competition (q^*,p^*), monopolistic pricing creates consumer surplus (CR), producer surplus (PR) and a dead-weight-loss (dwl) due to the fact that consumers willing to pay a price (p) $p^* \leq p \leq p_m$ are left empty-handed. As a matter of fact, the incentive created by granting a (temporary) monopoly comes at a cost as monopolistic pricing typically induces a dead-weight-loss and hence (static) allocative inefficiency. In other words, an optimal patent length (t* = T)[10], from the perspective of the society, is defined by a t* that maximises social welfare (SW), where[11]

$$(1)\ SW = \sum_{t=0}^{\infty}(CR_t + PR_t + dwl_t) \cdot e^{-rt} - \sum_{t=0}^{T} dwl_t \cdot e^{-rt} - s(t,R) \cdot R.$$

The first part looks at the total (possible) welfare from the new market for the product innovation, discounted by a factor e^{-rt}, which takes into account that future net benefits diminish over time (t) by an interest rate (r).[12] The second part of the equation reveals that for the time until the patent expires (in time t = T), the monopolisation effect of the patent induces a dead-weight-loss, also discounted by a factor e^{-rt}. Finally, the last part ($s(t,R)\cdot R$) reveals the social cost of the innovation, where the effort invested by a company (i.e. R&D-input R) is concave, pointing to an increasing but diminishing marginal return from investing in the production of an innovation, and s(t,R) being the variable cost. These variable costs have two components: first, the patent holder has to pay a patent fee that is increasing progressively over time (t). Second, the patent holder pays the variable cost for each unit of R. Moreover, this last part also reveals a necessary condition for the creator's incentive to create the innovation, namely the financial gains must be above the private investments, i.e. $\sum_{t=0}^{T} PR_t \cdot e^{-rt} \geq s(t,R) \cdot R$. Accordingly, the optimal design of a patent (or an IPR in general) can be seen as a balancing of static (and dynamic) cost versus benefits, which automatically leads to feature number two of the economics of IPR. We hence find good reasons to grant an IPR for a limited duration of time (T = t*),[13] as the implementation of excludability always comes at a cost.

Secondly, the economics of IPRs also emphasise the importance of the information function, i.e. the fact that the new intellectual property must be published, highlighting the dynamic character of the development of innovations as a cumulative production process and hence directly accounting for possible dynamic costs. The innovator does not only raise prices for existing products (static efficiency), but also for future products (dynamic efficiency) as the innovation path typically evolves cumulatively, i.e. basic innovations are being developed to further create new

[10] Please note that there are two dimensions defining the optimal design of a patent. While patent length takes into account the time during which the innovator is granted a monopoly, patent breadth defines the technology that is protected against imitation. See Posner (2005) on the general economics of IPR.

[11] We use areas (PR, CR and dwl) for simplicity reasons. Obviously, the rents internalised by the parties involved stem from the integral using the profit function ($\pi(q) = R(q) - C(q)$, i.e. profit is returns minus costs) of the company.

[12] See Nordhaus (1969) on the theory of optimal patent length. See e.g. Gallini (1992), Maurer/Scotchmer (1998) and Gilbert/Shapiro (1990) on the optimal mix of patent length and breadth. For a literature review, see Leveque/Meniere (2004).

[13] Note that not all types of IPR are granted for a limited duration of time, as trademark law is granted for a duration of 10 years with an infinite option of extension (i.e. infinite duration). However, in contrast to patents (with a maximum of 20 years) and copyright (with a duration of lifetime plus 70 years), trademark law seeks to correct another form of market failure. While patents and copyright seek to overcome the incentives problem due to the public good nature of the information good, trademark law is directed towards solving a problem of asymmetric information in the context of an adverse selection (Scheufen, 2018, 255).

products and/or applications. Therefore, granting a patent (and hence a monopolistic right) comes with the obligation to publish the information underlying the innovation in a patent document within 18 months after the application date.

The patent document shall create the absorptive capacity (Cohen/Levinthal, 1989) to enable an advancement of existing innovations, directly accounting for the dynamic character of the innovation process. We hence find that the economics of IPRs highlight the incentive and information function as the general intuition of an IPR system. However, consciousness of the second-best character of an IPR for public goods is crucial, as one market failure (externalities due to an undefined property right) is corrected by creating another market failure (market power due to a (temporary) monopolistic right). Extending on our market failure argumentation and the insights gained from the economics of IPR, the question arises as to whether an IPR for data is an efficient answer to the externality problem, and what the optimal design for such an IPR should feature.

4.1.2 Is There a Need for a New IPR for Data?

The fact that the amount of existing, and in some instances open, data has been increasing exponentially throughout the last decade, may at least lead us to question the need of an IPR for data from an incentives point of view. Obviously, there seem to be other motives for data creators to put effort into the creation of non-personal data. Moreover, the bulk of machine data basically creates itself, with virtually no data production costs (section 3.2.1). Nevertheless, as previously discussed, there may be specific forms or types of data which are a public good, and for which an incentives problem may therefore be predominant. As shown, for this type of data we may observe substantial data creation costs, revealing that the creator needs an incentive to make this large investment. The reason being that, without financial gains from collecting and trading data, the innovator will not be able to recoup his investment. To date, there has been no explicit proposal outlining the characteristics of a new (intellectual) property regime for data. However, in the European debate at least, there is consensus that such a new IPR could only be introduced through new legislation (Zech, 2016; Dorner, 2014).[14] With respect to the specific design or the features of such a new IPR, we find several recommendations in the legal and economic literature. We have pointed to four general features that an intellectual property right for data will have to encompass: subjectivity, totality, exploitability and the scope of protection (section 2.2.2).

- Subjectivity

  From an economic standpoint, the incentives argument may in fact only be valid for the producer of data. Accordingly, if we agree that an IPR is needed to stimulate large investments in data production, only the producer himself will have to be addressed by creating the expectation of future financial gains from his investment. However, and as stated before,

[14] See also Kerber (2016) on page 991.

it is not always clear who the producer of such data is, i.e. in some instances it may be questionable to whom the IPR would have to be assigned.

Moreover, for the majority of data, especially automated data, the investment in data production is close to zero. Thus, the only valid argument (i.e. we need an incentive for the creator of data to stimulate data creation) to incorporate subjectivity or ownership for data is not valid for the majority of data.

- Totality

  From an economic point of view, the totality feature may be difficult to ensure in the context of data for at least two reasons. Firstly, even an intellectual property right may not prevent somebody else from lawfully collecting the very same data on his own. In fact, priority as the predominant feature in the patent system may not be applicable in the context of data, but rather a system similar to the copyright regime, where no priority is given to the producer of a copyrightable work. As stated before, data markets may also be characterised partly by large network effects and natural monopolies. One may at least question whether an IPR would enhance such network externalities or not. Secondly, the enforcement of an IPR for data would at least be very limited for published data, the reason being that data can be easily copied and distributed once it has been published. As a result, totality may only be guaranteed for unpublished data, which strongly questions the need of an IPR for data in the first place.

- Exploitability

  With respect to data, we doubt that a right is needed to ensure that data is used and traded. However, it is very unclear whether an IPR is the only and best mechanism to ensure exploitability. In fact, there are good reasons to believe that an IPR for data may even exacerbate problems of access to data and judicial interventions in the form of compulsory licensing, likely to boost the costs of the judicial system. As a result, an IPR may not be a second best (recalling that a first-best solution cannot be reached by granting market power), but rather a third- or fourth-best solution for correcting an unlikely incentives problem, and hence missing exploitability in the context of data.

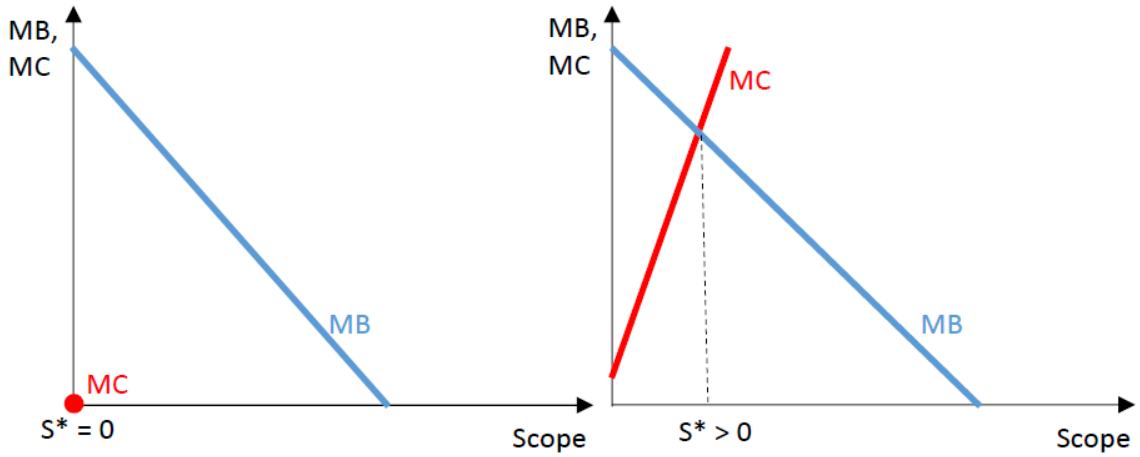
- Scope of Protection

  Lastly, the scope of protection sets out the range of protection. Accordingly, for patents we differentiate between patent length (usually 20 years), breadth (whether or not similar inventions may compete, i.e. the cost of an inventing around)[15] and depth (priority to discovery, i.e. no double creation). For copyright protection, we distinguish between copyright length (usually author lifetime plus 70 years), breadth (exceptions of copyright) and depth (no priority, i.e. double creation is possible).

[15] There are different ways of defining patent breadth. The cost of an inventing around determine the cost to create a similar technology that differs just enough to be patentable itself. See particularly the pioneering works of Gallini (1992), Gilbert/Shapiro (1990) and Maurer/Scotchmer (1998).

From an economic standpoint, there are several good reasons why the scope of such a new IPR should be very limited with respect to the scope of protection. In fact, the intersection between marginal benefits and marginal costs of additional scope of protection may be in the zero range (figure 4-2). On the one hand, we find that the marginal benefits of data decline with additional scope and there is no reason to believe that this would differ when comparing automated versus non-automated data. The reason for the downward sloping marginal benefit function is simply that the future returns from an additional year of protection (scope) decrease the longer the returns accrue over time. This can also be seen from the social welfare function in section 4.1.1, as every future net return is discounted by the term e^{-rt}. On the other hand, the marginal cost may be completely independent of scope for automated data, as the marginal costs for production are always zero. For non-automated data, in contrast, we argue that the marginal costs increase with the scope of protection, as the investment in data production may be substantial and increase over time. The reason the marginal cost increases over time is twofold: firstly, we have seen that the patent fee increases progressively, that is $\partial s(t,R)R/\partial t > 0$ in equation 1. Secondly, recalling the cumulative character of the innovation process each additional year of protection (or unit of scope more generally) adds licensing costs and/or litigation costs for advancements of protected basic innovations. Moreover, given that the production costs for non-automated data can be substantial, the marginal cost function will not start in the zero point of our diagram (Figure 4-2 (b)). Thus, we may find two possible scenarios when discussing the optimal scope of protection.

Figure 4-2: Optimal Scope of an IPR for Data: Automated vs. Non-Automated Data



(a) Scope for Automated Data (b) Scope for Non-Automated Data

Source: own depiction

Accordingly, there are good reasons to argue that for the majority of data (i.e. automated data), there is no optimal scope for a new IPR in terms of length, breadth and depth, as marginal cost and marginal benefits do no intersect.

As a result, social welfare – i.e. the difference between marginal benefits and marginal costs – will only be maximised if, and only if, the scope of protection is set to zero (S* = 0 in figure 4-2 (a)). Thus, there is simply no reason to introduce an IPR for automated data. For non-automated data on the other hand, the scope of protection should be limited to a very low level of protection (S* > 0 in figure 4-2 (b)). In this regard, Zech (2016) favours a short-term protection, which may be adequate for non-automated data only, as figure 4-2 reveals. With respect to the depth of protection, we have already pointed to the fact that a lawful and independent reproduction of identical data sets should be possible, i.e. the depth of protection should be extremely limited. Lastly, the breadth of protection highlights a characteristic that differs considerably when comparing data with other information goods. In the context of data, a matching of different forms and origins of data sets is a prerequisite for enabling the internalisation of decisive welfare gains. As such, the breadth of protection should be as narrow as possible to avoid limited access preventing such matching processes.

To summarise, despite some good arguments in favour of a new IPR for specific forms of data, we should be very reluctant to introduce such a regime, as the majority of data are not defined by an incentives problem. Moreover, there are side effects that such a new property right in data would cause for the full set of data that would have been created despite a property right. Arguably, we have to bear in mind that the introduction of a(n) (intellectual) property right can never be a first-best option, since the market power created by granting a temporary monopoly is tantamount to a market failure. We have pointed to arguments as to why an IPR may not even reach a second-best but rather third- or fourth-best level in solving a market failure from not having well-defined property rights (assumption number 1 in Figure 3-1). In fact, there may be other forms of law and policy options that deal with the actual problem at hand, i.e. an access right to data.

4.2 Market Power: Antitrust Law as a Solution

As we have previously seen, a vast majority of market failure problems in the market for data may stem from market power as well as network effects, raising the issue of access to data. In fact, the data economy may reveal substantial social welfare gains only with respect to matching different data sets, benefitting from completely new spillover and additional information. For instance, to fight crime, it may not be sufficient to simply increase police presence on the streets. Instead, it could be effective to match regional crime statistics to reveal probabilities for specific regions on specific dates, and then increase police presence only selectively. However, in order to internalise such synergies in matching different data sets, sufficient access to data is necessary. While a new (intellectual) property right would even exacerbate the access problem, other forms of law, particularly competition law, but also contractual solutions, may (under certain circumstances) be able to solve externality, market power, as well as information asymmetry issues. A substantial reform of the existing regulations and legal norms, however, may be inevitable in this respect. Nevertheless, this change in regulation has already started. For example, the German government introduced a dominant position due to data in the last (9th overall) amendment of the German act against restraints on competition (Rusche, 2017). Accordingly, the relevance of data as an input for economic success in a digitalised economy was recognised

by the authorities. In light of this, the question arises as to whether a company with a dominant position due to data should also grant access to its data to guarantee working competition in the market within which it is active, or in other more or less related markets.

On the one hand, "an undertaking which holds a dominant position has a special responsibility not to allow its conduct to impair genuine undistorted competition in the internal market" (Court of Justice of the European Union, 2011, para. 24). Accordingly, a dominant company can generally be forced to grant access to data or information.[16] On the other hand, two preconditions must be fulfilled for forcing a dominant company to grant access (Bundeskartellamt, 2017, 7):

- the access to the requested data has to be important for economic success in a market and
- competing or not competing companies in the same or different markets are unable to buy or collect a set of data that is similar, or at least as useful (natural monopoly).

However, both preconditions are generally not fulfilled at the same time (Demary/Rusche, 2018, 52 ff.). In essence, the provision of data by using antitrust regulation is not very likely because for every single case, a company has to be deemed dominant because of data, it has to be proven that the data is necessary for success and that there is no other source of suitable data. Furthermore, the long duration of proceedings also plays a role, especially in the dynamic digital economy. For example, the proceedings in the Microsoft case ran from 2004 to 2007 (Court of Justice of the European Union, 2007).

However, antitrust regulation can also forbid a voluntary exchange of, and the access to, data. This is especially the case when a company already has market power thanks to data, and new data would reinforce its dominant market position. The exchange of data containing strategic information or from which strategic information can be deduced is also forbidden (Frenz, 2016, 673). Strategic information in this sense is any piece of information that is not publically available and (potentially) changes the conduct of a competitor, or reveals the company's own conduct to competitors. Examples may be upcoming price increases, bids in a public tender, or product specifications that can be deduced from automated data.

In essence, antitrust regulation can be used to gain access to the data of a dominant company. However, the preconditions that must be fulfilled and the long duration of proceedings make it rather inappropriate as a general regulation of access to data. Furthermore, the scope of data that can be exchanged voluntarily or by decree is limited on data that does not contain any strategic information or lead to market power.

4.3 Information Asymmetry: Contract Law as a Solution

[16] That this is not a mere theoretical result was shown by the European Court of Justice in its Judgement in the Microsoft case (Court of Justice of the European Union, 2007). Microsoft was convicted to grant access to information on compatibility to foster competition in bordering markets.

Complementary to a comprehensive competition policy approach, contractual solutions may also be able to solve some predominant problems associated with access to data. In this regard, bilateral agreements between companies as well as multilateral agreements similar to patent pools, i.e. a new form of data pool, could solve access issues to create the synergies from matching different data sources from a range of data holders. Nevertheless, technical features could also be able to determine whether such cooperation would work. A typical example is data that is generated in the context of connected cars, which allows for a number of additional services for the driver of a car, such as Advanced Driver Assistance Systems. In such settings, it may simply be the storage location for the data that defines who has access and who does not.

In fact, car manufacturers may have an incentive to prevent other service providers (e.g. component suppliers, insurers, public authorities, mobility service providers, (independent) aftermarket suppliers) from gaining access to the data, even though data sharing would allow for substantial (and especially complementary individual) welfare gains. In such cases, only regulatory norms are able to enhance the internalisation of such gains via compulsory data sharing. For settings in which companies seek data sharing, however, technical barriers could prevent free and efficient sharing and matching. In this context, specific platforms such as the Industrial Data Space (IDS)[17] may enhance contractual solutions in forms of bilateral and multilateral agreements.

The IDS is a sovereign data architecture that companies can use for trading and/or sharing data. Most interestingly, the platform allows for the clear definition and securing of specific forms of data usage, creating a technological baseline for implementing data sharing based on bi- or multilateral agreements or data pools. Of course, such cooperation reveals potentials for collusive behaviour,[18] greatly highlighting the need for regulatory and third-party oversight. With respect to antitrust regulation and an industrial data space, it must be said that whether the strategic information and data from which such information can be deduced is exchanged directly or indirectly is irrelevant (Court of Justice of the European Union, 2016a, para. 27). This was underlined by the European Court of Justice in the so called “Eturas” case (Court of Justice of the European Union, 2016a), which involved a travel agency that ran a platform on which other agencies could sell their services. The administrator sent an electronic message to at least some of the travel agencies that were active on the platform. The aim of the message was for the various agencies involved to coordinate and thus limit discounts on the products sold. The judges found that this was a concerted practice and therefore forbidden according to antitrust regulation.
In this case, however, it was also underlined that merely dispatching a message is generally not sufficient to be considered part of a concerted practice. Further evidence, such as a change in behaviour, i.e. a reduction in the discount applied, is necessary (ibid, para 50).

[17] Since 2018, the term “international data space” is also commonly used to describe the international character and scope of the concept. See e.g. Fraunhofer ISST (2018) for an overview on the IDS.

[18] In this respect, Art. 101 (3) AEUV clarifies the exceptions of the prohibition of cartels. Accordingly, a data pool may state an exception if, and only if, such cooperation is necessary to facilitate technological or economic progress and is limited in scope.

In relation to data, this implies that a company exchanging data in an industrial data space has to take into account which data is exchanged with whom and, which data are made available in an industrial data space. Moreover, data that is acquired must be checked for strategic information.

5 Conclusions and Outlook

In conclusion, we find that an (new) intellectual property right for syntactic data is, overall, neither a feasible nor an efficient solution for solving the various market failures involved in the market for non-personal data. In fact, we have pointed to several specific properties (in the context of subjectivity, totality, exploitability and the scope of protection) that may prevent assigning ownership to a right holder, as well as causing other side effects arising from introducing a new IPR regime. We have also shown that the incentives problem may only be a problem for few specific sectors, characterised by substantial investments for data collection, particularly non-automated data. For the vast majority of automated data, we support the positions stated in the literature and by stakeholders in the economy, whereby the incentives problem is not considered the most significant market failure in the data market. Instead, it is the issues surrounding the right to access data that will have to be resolved, by finding competition policy and contract law solutions to stimulate sharing and matching of data to enable internalisation of the potential welfare gains in the data economy. It is also dubious, in this respect, whether a new IPR will be able to adjust to the developments in the dynamic digital economy or if it would hamper the use and sharing of data.

We have shown that there are already various options for defining data ownership by referring to the Database Directive or trade secret law. Additional property rights may in fact even exacerbate the market outcome, inevitably leaving us with an inefficient allocation of data resources in terms of, at most, a third- or fourth-best solution. The recent competition law jurisprudence has already taken notable steps to foster the competitiveness of new business models in the data economy and to prevent dominant market positions of companies in the data market. Nevertheless, developments in the digital economy can also make the introduction of a new IPR worthwhile in the near future. Furthermore, we have also pointed to some specific characteristics for which the policymaker will have to find feasible and adequate answers when the internet of things and future business models for trading data proceed. In fact, the various approaches in the literature that try to define ownership of syntactic data reveal that legal uncertainty exists, which may also lead to a fragmentation of the European Single Market because of the use of differing approaches in different member countries.

References

Assion, Simon / Heun, Sven-Erik, 2015, Internet(recht) der Dinge. Zum Aufeinandertreffen von Sachen- und Informationsrecht, in: Computer und Recht, Vol. 31, No. 12, pp. 812–818

Buchanan, James / Yong, Yoon, 2000, Symmetric Tragedies: Commons and Anticommons, in: Journal of Law and Economics, Vol. 43, No. 1, pp. 1–13

Bundeskartellamt, 2017, Big Data und Wettbewerb. Schriftenreihe „Wettbewerb und Verbraucherschutz in der digitalen Wirtschaft“, No. 1, Bonn

Bundesministerium für Verkehr und digitale Infrastruktur. 2017, Wir brauchen ein Datengesetz in Deutschland, Strategiepapier Digitale Souveränität, https://www.bmvi.de/SharedDocs/DE/Artikel/DG/datengesetz.html [25.09.2018]

Bundesverband der Deutschen Industrie, 2015, Industrie 4.0 - Rechtliche Herausforderungen der Digitalisierung. Ein Beitrag zum politischen Diskurs, Berlin, München

Coase, Ronald, 1960, The Problem of Social Cost, in: Journal of Law and Economics, Vol. 3, pp. 1–44

Cohen, Wesley M. / Levinthal, Daniel A., 1989, Innovation and learning: The two faces of r&d, in: Economic Journal, in: Economic Journal, Vol. 99, No. 397, pp. 569–596

Court of Justice of the European Union, 2007, Judgment in Case T-201/04, Microsoft Corp. v European Commission

Court of Justice of the European Union, 2011, Judgment in Case C-52/09, Konkurrensverket v TeliaSonera Sverige AB

Court of Justice of the European Union, 2016a, Judgment in Case C-74/14, Eturas, http://curia.europa.eu/juris/document/document.jsf?text=&docid=173680&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=1396676 [31.10.2018]

Court of Justice of the European Union, 2016b, Judgment in Case C-582/14, Patrick Breyer v. Bundesrepublik Deutschland, Luxembourg, http://curia.europa.eu/juris/document/document.jsf;jsessionid=9ea7d0f130dcb550c147a1fb41d1b647b9a3d7036bd1.e34KaxiLc3eQc40LaxqMbN4Pbh0Ne0?text=&docid=184668&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=693310 [31.10.2018]

Demary, Vera / Rusche, Christian, 2018, Economics of Platforms, IW-Analyse, No. 123, Cologne

Demsetz, Harold, 1967, Toward a Theory of Property Rights, in: American Economic Review, Vol. 57, No. 2, pp. 347–359

Dorner, Michael, 2014, Big Data und „Dateneigentum“, in: Computer und Recht, Vol. 30, No. 9, pp. 617-628

Drechsler, Jörg / Jentzsch, Nicola, 2018, Synthetische Daten. Innovationspotential und gesellschaftliche Herausforderungen, Impuls paper, https://www.stiftung-nv.de/sites/default/files/synthetische_daten.pdf [04.09.2018]

Drexl, Josepf, 2017, Designing Competitive Markets for Industrial Data. Between Propertisation and Access, in: Journal of Intellectual Property, Information Technology and Electronic Commerce Law, Vol. 8, No. 4, pp. 257–292

Duch-Brown, Nestor / Martens, Bertin / Mueller-Langer, Frank, 2017, The Economics of Ownership, Access and Trade in Digital Data. Digital Economy Working Paper 2017-01, Seville

Duisberg, Alexander, 2017, "Datenhoheit und Recht des Datenbankenherstellers" - Recht am Einzeldatum vs. Rechte an Datensammlungen, in: Daten als Wirtschaftsgut, pp. 15–35

Eger, Thomas / Scheufen, Marc (2018): The economics of open access. On the future of academic publishing, Cheltenham

Ensthaler, Jürgen, 2016, Industrie 4.0 und die Berechtigung an Daten, in: Neue Juristische Wochenzeitschrift, No. 48, pp. 3473–3478

European Commission, 2015, A Digital Single Market Strategy for Europe. Com(2015) 192 final, Brüssel, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52015DC0192&from=EN [1.10.2018]

Franceschi, Alberto De / Lehmann, Michael, 2015, Data as Tradeable Commodity and New Measures for their Protection, in: The Italian Law Journal, Vol. 1, No. 1, pp. 51–72

Fraunhofer ISST, 2018, Industrial Data Space. Überblick, Köln

Frenz, Walter, 2016, Industrie 4.0 und Wettbewerbsrecht, in: Wettbewerb in Recht und Praxis, Vol. 62, No. 6, pp. 671–678

Fritsch, Michael, 2014, Marktversagen und Wirtschaftspolitik. Mikroökonomische Grundlagen staatlichen Handelns, München

Gallini, Nancy, 1992, Patent Policy and Costly Imitation, in: RAND Journal of Economics, Vol. 23, No. 1, pp. 52–63

Gilbert, Richard / Shapiro, Carl, 1990, Optimal Patent Length and Breadth, in: RAND Journal of Economics, Vol. 21, pp. 106–112

Grosskopf, Lambert, 2012, Rechte an privat erhobenen Geo- und Telemetriedaten, in: Strobl, Josef / Blaschke, Thomas / Griesebner, Gerald (ed.), Angewandte Geoinformatik 2012. Beiträge zum 24. AGIT-Symposium Salzburg, Berlin, pp. 171–174

Haucap, Justus, 2015, Ordnungspolitik und Kartellrecht im Zeitalter der Digitalisierung, DICE Ordnungspolitische Perspektiven No. 77, Düsseldorf

Haucap, Justus / Kerber, Wolfgang / Schweitzer, Heike / Welker, Robert, 2018, Modernisierung der Missbrauchsaufsicht für marktmächtige Unternehmen, Endbericht Projekt im Auftrag des Bundesministeriums für Wirtschaft und Energie (BMWi), Projekt-No. 66/17, o.O.

Heckman, Judd / Peters, Elizabeth / Kurup, Nikhil Gopinath / Boehmer, Erin / Davaloo, Milad, 2015, A Pricing Model for Data Markets, Newport Beach

Heller, Michael A., 1998, The Tragedy of the Anticommons: Property in the Transition from Marx to Markets, in: Harvard Law Review, Vol. 111, No. 3, p.621

Hennemann, Moritz / Paal, Boris, 2018, Big Data As An Asset. Daten und Kartellrecht, Gutachten im Projekt ABIDA – Assessing Big Data, o.O.

Hoeren, Thomas, 2013, Dateneigentum. Versuch einer Anwendung von § 303a StGB im Zivilrecht, in: MultiMedia und Recht, Vol. 16, No. 8, pp. 486–491

Hugenholtz, Bernd, 2017, Data Property in the System of Intellectual Property Law: Welcome Guest or Misfit, in: Lohsse, Sebastian / Staudenmayer, Dirk / Schulze, Reiner (ed.), Trading Data in the Digital Economy: Legal Concepts and Tools. Münster Colloquia on EU Law and the Digital Economy III, Baden-Baden, pp. 75-100

IDC - International Data Corporation and Open Evidence, 2017, European Data Market, SMART 2013/0063, Final Report, Study for the European Commission (Directorate-General for Communications Networks, Content and Technology), Brussels

Jänich, Volker, 2002, Geistiges Eigentum – eine Komplementärerscheinung zum Sacheigentum?, Tübingen

Kerber, Wolfgang, 2016, A New (Intellectual) Property Right for Non-Personal Data? An Economic Analysis, in: Gewerblicher Rechtsschutz und Urheberrechte, Internationaler Teil (GRUR Int), Vol. 65, No. 11, pp. 989–999

Kiefer, Jonas, 2018, Das Geschäftsgeheimnis nach dem Referentenentwurf zum Geschäftsgeheimnisgesetz: Ein Immaterialgüterrecht, in: Wettbewerb in Recht und Praxis, Vol. 64, No. 8, pp. 910–917

Leveque, Francois / Meniere, Yann, 2004, The Economics of Patents and Copyright, Paris

Lohsee, Sebastian / Schulze, Reiner / Staudenmayer, Dirk, 2017, Introduction, in: Lohsse, Sebastian / Staudenmayer, Dirk / Schulze, Reiner (ed.), Trading Data in the Digital Economy: Legal Concepts and Tools. Münster Colloquia on EU Law and the Digital Economy III, Baden-Baden, pp. 13–24

Maurer, Stephen / Scotchmer, Suzanne, 1998, The Independent Invention Defense in Intellectual Property, in: Economica, Vol. 69, No. 276, pp. 535–547

Nordhaus, William, 1969, An Economic Theory of Technological Change, in: The American Economic Review, Vol. 59, No. 2, pp. 18–28

Official Journal of the European Communities, 1996, Directive 96/9/EC on the legal protection of databases, Vol. 39, L 77, pp. 20–28

Official Journal of the European Union, 2016a, Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), Vol. 59, L 119, pp. 1–88

Official Journal of the European Union, 2016b, Directive (EU) 2016/943 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure, Vol. 59, L 157, pp. 1–18

Ohly, Ansgar, 2014, Der Geheimnisschutz in deutschen Recht. Heutiger Stand und Perspektiven, in: Gewerblicher Rechtsschutz und Urheberrechte, Vol. 116, No. 1, pp. 1–11

Plattform Industrie 4.0, 2016, Industrie 4.0 - Wie das Recht Schritt hält, Berlin

Posner, Richard A., 2005, Intellectual Property: The Law and Economics Approach, in: Journal of Economic Perspectives, Vol. 19, No. 2, pp. 57–73

Rifkin, Jeremy (2015): The zero marginal cost society. The internet of things, the collaborative commons, and the eclipse of capitalism, New York

Rusche, Christian, 2017, Wettbewerbsrecht 2.0, IW-Kurzbericht, No. 20, Cologne

Rusche, Christian, 2018 Forthcoming, Data economy and antitrust regulation. An European perspective, in: Intereconomics

Scheufen, Marc, 2015, Copyright versus open access. On the organisation and international political economy of access to scientific knowledge, Cham

Scheufen, Marc, 2018, Angewandte Mikroökonomie und Wirtschaftspolitik. Mit einer Einführung in die ökonomische Analyse des Rechts, Berlin, Heidelberg

Varian, Hal R., 2003, Intermediate Microeconomics. A modern Approach, New York, London

Weber, Robert, 2017, Improvement of Data Economy Through Compulsory Licenses?, in: Lohsse, Sebastian / Staudenmayer, Dirk / Schulze, Reiner (ed.), Trading Data in the Digital Economy: Legal Concepts and Tools. Münster Colloquia on EU Law and the Digital Economy III, Baden-Baden, pp. 137–158

Zech, Herbert, 2015, Daten als Wirtschaftsgut – Überlegungen zu einem „Recht des Datenerzeugers“, in: Computer und Recht, Vol. 31, No. 3, pp. 137–146

Zech, Herbert, 2016, A legal framework for a data economy in the European Digital Single Market: rights to use data, in: Journal of Intellectual Property Law & Practice, Vol. 11, No. 6, pp. 460–470