JRC TECHNICAL REPORTS

JRC Digital Economy Working Paper 2017-01

The economics of ownership, access and trade in digital data

Nestor Duch-Brown
Bertin Martens
Frank Mueller-Langer

2017

Table of Contents

Abstract

Despite the rapidly growing volume and economic importance of data in the digital economy, the legal framework for data ownership, access and trade remains incompletely defined in the EU and elsewhere. De facto data ownership dominates and often leads to fragmentation or anti-commons problems in data. Combined with limited access and trade, this inhibits the realisation of the full economic benefits of non-rival data. It may slow down innovation and affect the efficiency of data markets. We examine three potential sources of data market failures: externalities related to economies of scope in data, strategic behaviour of data owners and transaction costs in data exchanges. We link the legal debate on data ownership with relevant branches of the economics literature, including intellectual property rights economics, the commons and anti-commons literature, models of trade under the Arrow Information Paradox and multi-sided markets. Economists are inclined to think that well-defined private property rights are a necessary condition for an efficient resource allocation. The question in this paper is to what extent this view holds for non-rival data. We show that the allocation of data ownership or residual control rights matters, not only for private benefits but also for social welfare. The outcomes of bargaining over data ownership and access rights do not necessarily maximize social welfare. Can regulators intervene to improve these outcomes? Would a better specification of legal ownership rights or introducing access provisions to improve efficiency and reduce data market failures? There are no easy answers to these largely empirical questions. We offer no policy solutions yet and more research is required to bring economics up to speed with these questions.

1.   Introduction

Digital information technology has lowered the cost of collecting, distributing and using data to reduce search costs. Very large online markets offer a huge variety of products at very competitive prices and low transaction costs. Digital algorithms help to overcome the problem of information friction in these markets with heterogeneous products and user preferences. Market entry and exit costs have declined to the point where direct peer-to-peer exchanges in collaborative platforms become feasible. Many firms are exploring and experimenting with these technologies. They can potentially generate many benefits for producers and users. Digital technology has made behaviour more easily observable at low information costs, raising questions about the protection of personal and commercial data. The "observer effect" (also referred to as Hawthorne effect)[1] may come into play as subjects start adjusting their behaviour knowing that they might be observed.

Uncertainties about the impact of these digital data technologies on human behaviour, markets and social welfare are a source of concern for citizens and firms and triggers regulatory questions for policy makers, in particular questions about ownership, access and trade in digital data and potential data market failures that may require regulatory intervention are gaining importance. The legal and regulatory framework for data ownership, access and trade has not kept pace with technological advances. It remains vague and incomplete both in the EU and elsewhere. The European Commission's Digital Single Market (2015) policy package aims to address at least some of these concerns. It launched the European Data Economy initiative as part of its Digital Single Market policy agenda and a public consultation on these issues in January 2017. The OECD (2015, 2016) and several EU Member States are also addressing these issues, for example the Loi Lemaire in France.

Economic thinking on the role of data in the digital economy is also lagging behind advances in data technology. While economists are very active on digital economy issues they have done little work on the economics of data ownership, access and trade in data markets and remain on the sideline in the policy debates. To their credit, it must be said that they have come a long way to adapt to the information age. The previous century saw the rise and fall of the neo-classical economic welfare paradigm that was based, amongst others, on the unrealistic assumption of perfectly available information at zero cost – long before digital information technology was invented. The work of Stiglitz (2000) and several others showed that this paradigm fails under conditions of imperfect information and that access to and the distribution of information matters for market outcomes and welfare. Today, positive information costs and asymmetrically distributed information are part of mainstream economics. Digital technology lowers information costs but does not solve information asymmetries. On the contrary, as the volume of data and information grows exponentially, it may exacerbate it and affect the behaviour and welfare of firms and individuals. With digitization, information ownership, access and trade may even matter more for economic welfare outcomes.

The objective of this report is to provide an overview of the rather scarce economic research literature on data ownership, access and trade. We start with defining some basic concepts such as data and information and the technical properties of digital data that determine their economic

---

[1] For instance, see Landsberger (1958), Monahan and Fisher (2010) and Levitt and List (2011).

characteristics (Section 2). We inquire how the economic characteristics of data differ from ordinary goods. Section 3 briefly explores the legal perspective on data ownership and access. We conclude that data ownership is only weakly and incompletely defined. De facto ownership seems to dominate the data economy. Some competition scholars argue that access to data is not a problem for the efficiency of data markets since data are substitutable; others argue the opposite.

Economists are generally inclined to think that well-specified property rights reduce transaction costs and uncertainty and thereby increase the efficiency of markets. How do incompletely defined data ownership rights affect market outcomes and economic welfare and how can this situation be improved, if at all? Translated into economic jargon: What are the potential sources of data market failures and can regulatory intervention, for instance through a more complete definition of data ownership rights, correct these failures? In this paper we examine three potential sources of market failures: externalities, strategic behaviour by data owners and transaction costs.

We start exploring some possible answers to these questions through the lens of the economics of intellectual property rights (Section 4.1). This approach connects well with the legal debate in the EU where the Database Directive builds on the foundations of copyright as an incentive to invest in data production and trade. We show how the economic characteristics of data generate strong externalities that make copyright-like protection less appropriate for data (Section 4.2). Externalities provide arguments for weaker data ownership rights and stronger access rights for potential data users. Another well-known source of market failure in intellectual property rights (IPRs) is strategic behaviour by rights holders. Section 4.3 uses a number of examples from digital markets and applies some simple concepts from game theory to explain how strategic behaviour leads to under-utilisation of data and reduces the potential social welfare that can be gained from data sources.

Section 5 presents some features of existing data markets and explores the economic literature on data trade, including the transaction cost issues associated with these markets. Some economists assume that data ownership and access are sufficiently well-defined and focus on the characteristics of trade in data. That starts from the Arrow Information Paradox: once data are revealed they cannot be traded anymore because nobody pays a price for known data. In practice however there are many ways to differentiate data and price-discriminate between buyers. Other researchers examine the role of data in multi-sided product markets or platforms, how data facilitate transactions and matching of users on different sides of the market, for example through ad auctions and search rankings. There are concerns about the trade-off between efficiency and revenue in such markets. Finally, Section 6 discusses data access, portability and interoperability.

We conclude that there are no easy answers for regulators how to overcome market failures in data and information markets. This paper is a call for more research on these questions.

2.   Defining data and their economic properties

It is important to understand what we are talking about. In everyday parlance we often use several words interchangeably: data, information, communication, etc. However, they refer to different things. The purpose of this section is not only to define these objects but also to describe some of their key technical characteristics and how they affect the economic impact of digital technology. That will enable us to get a better understanding of the economic implications of the legal aspects of data ownership, access and trade.

2.1.   What are data, information and communication?

There are many views and theories of information. While there is no full consensus there are some basic elements where a consensus seems to emerge. Information theory starts with applications in telecommunications by Hartley (1928) and Shannon (1948). Their engineering approach focuses on the transmission of signals or messages between a sender and a receiver, irrespective of the semantic meaning of the message. All transmission requires a physical carrier; there are no non-embodied messages: electrons for electronic messages, photons for visual messages, sound waves at supra-molecular level for spoken messages, paper for written messages, molecules for chemical messaging systems, etc. Information theory defines information as the number of discernible signals or data points needed to transmit a message (Boisot & Canals, 2004). For example, the number of binary bytes (0-1) or Morse signs (short-long) required to transmit a text message. Information contained in signals or messages exists prior to being detected by the receiver. The grass is green, independent of human observation.

Discernible signals in messages come in two formats: continuous and discrete. The former are called analogue; the latter digital. For example, when we speak we produce an analogue continuous sound wave that varies in frequency and amplitude. When we write down the spoken words we split them up in discrete symbols that represent these sounds and words. The Latin alphabet consists of 26 discernible symbols[2]. Hindu maths marked the emergence of a 10-digit calculus system that we still use today while the Romans kept struggling with an unwieldy 5-digit system. As such, human designed digital information systems have existed for thousands of years, ever since discrete symbolic communication was invented. Ants already had digital information systems, based on discrete chemical signals, for the last 150 million years (Holldobler & Wilson, 1990). Electronic signal systems can vary continuously in voltage level but in digital computers this analogue variation is artificially reduced to two discrete voltage steps that distinguish between 0 and 1 only. Binary counting rules and the extended ASCII system link packages of binary bits to specific symbols in the alphabet in order to enable text messages. Many software packages have their own translation rules

---

[2] Discernible states depend on scale. In an ordinary picture, the number of discernible shapes, colours, tones and sizes is the data content of the picture. One can of course examine the picture with a magnifying glass and identify each pixel. Most people only look with the naked eye and only distinguish between shapes and colours. That is usually the relevant data scale of a picture, unless you are a cryptographer who attempts to discover hidden keys in the pixel structure. Quantum scientists would go down to the subatomic level and may discover ambiguous quantum states. We are not concerned with that level of analysis.

from binary code to more aggregated interpretations. This translation is carried out by compiler programmes.

Humans are not very good at processing binary digital code and usually interact with electronic data through an interface. Nobody reads the binary code that represents the text of an e-book; we prefer to read the text in the 26-character alphabetic code, arranged in lines and pages on a screen or a paper print-out. We don't write binary code but type alphabetic characters on a keyboard that translates them into binary code. More generally, sensors are required to collect data from human activity (pictures taken through a lens, sound recorded by a microphone, text typed on a keyboard, touch screens on a mobile phone, etc.) or from natural events (temperature and wind, earth observations, physical, biological and chemical processes in nature and in industrial production, etc.) and to translate binary code into humanly observable output, via texts, screens, etc. In that sense, all digital data are machine-generated[3].

Some authors make a distinction between machine-generated and human-generated data. Zech (2016) defines "data" as machine-readable code created through automated measurement processes. Data are defined at the syntactic level. This does not include higher-level semantic content such as media content or software, and does not cover lower-level physical carriers of data. That excludes intellectual creations that are covered by copyright. The separation line between direct and indirect human data sources is not clear however. Many machine-generated data have a direct or indirect human source. For example car engine and house temperatures, electricity consumption meters etc., provide indirect information on human behavior. This debate comes back in the discussion on the distinction between personal and non-personal data (see Chapters 3 and 4). The Commission Communication on the Data Economy (2016, p. 5) defines machine-generated data as data "created without the direct intervention of a human by computer processes, applications or services, or by sensors processing information received from equipment, software or machinery, whether virtual or real". Machine-generated data as a concept is sometimes used to distinguish between personal and non-personal data. However, it is not very useful for that purpose either because it still leaves open the question of the borderline between direct and indirect human intervention. The recognition that all data are produced by machines and sensors built into machines points to the key data ownership issue that will be discussed in Chapter 4: the sensor/machine may be owned by another party than the data collector and the subject whose data are collected. The value of the data can only be realized if all parties collaborate. Allocating a data ownership right to one or the other party does not solve that issue.

The distinction between data, syntax and semantic content is crucial for understanding the importance of data analytics. Information is the semantic content that can be extracted from data or signals. Vigo (2013) and Boisot & Canals (2004) argue that extraction of the semantic content from a signal, a message or a data set requires a combination of prior structural knowledge (understanding the discrete symbols or the continuous signal, their semantics and syntax) and contextual knowledge, acquired through learning (see Figure 1). Vigo (2013) reformulates this extraction process as the number of discernible categorical data points needed to understand the

---

[3] See Wikipedia on the definition of "Machine-generated data". The terminology is also used in the European Commission Communication (2017) on the Data Economy.

meaning of a message, using categorisation theory from cognitive psychology. Categorisation can cover the symbols, semantics and syntax as well as the contextual meaning. For example, thanks to my knowledge of the alphabet I can read the words of a French text but without prior knowledge of French syntax and vocabulary I cannot understand the meaning of a French text. Even if I understand French I may not understand the meaning of a French text on quantum physics because I have no prior knowledge of that subject. The distinction between data and semantic content is also important to distinguish media products such as books, music, film and news articles from the underlying data that carry this content. Media content is protected by copyright, a well-known intellectual property rights system. Data are not protected by copyright (see Chapter 3).

Figure 1: The interaction between the World, data signals, information and knowledge
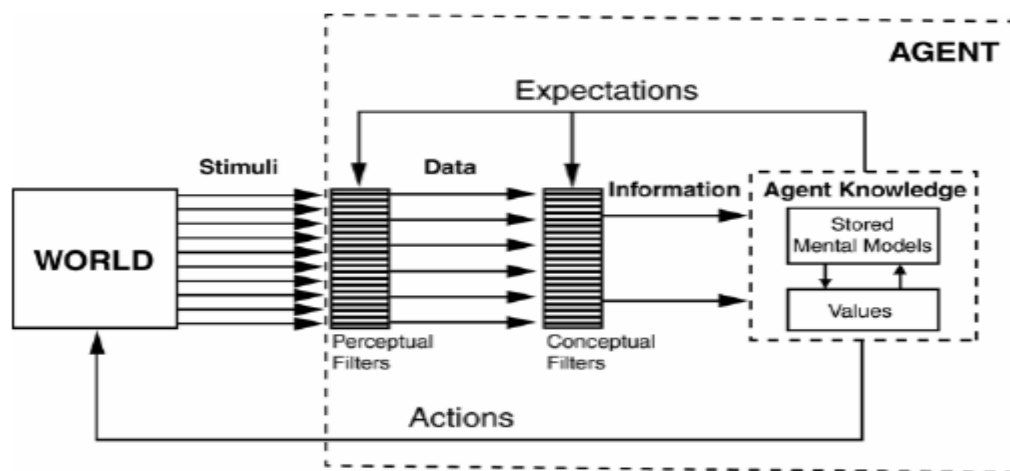
Source: Boisot & Canals, 2004.

Complexity theory (Gell-Mann, 1996) bridges the gap between information and knowledge. The more regularities can be observed in a seemingly chaotic dataset, the shorter the length of an exhaustive description of that dataset and the more complex the dataset is. The accumulation of knowledge through learning consists of identifying or extracting regularities from an apparently chaotic set of data, which can be used to predict data points and to guide behaviour. For example, learning a language consists of detecting the regularity of words in a set of apparently meaningless and chaotic sounds, linking these words to semantic meaning and understanding the grammatical regularities that link words with each other and give them different meanings. The knowledge contained in a dictionary and a grammar constitutes a concise and highly complex description of a larger set of alphabetic symbols and word strings. It can be used to guide our efforts at speaking the language.

2.2. Economies of scope in data analytics

Extracting content from a dataset or signal can be done at various levels of complexity. Kolmogorov (1963) defined the informational content of a message, such as a piece of text, as the length of the shortest computer program (in a predetermined programming language) that produces this text as

output. Complexity is a measure of knowledge or regularities extracted from a dataset. Regularities can be discovered through learning that contributes to acquiring knowledge. Knowledge can be used to predict data points and to guide behaviour. For example, learning a language consists of detecting the regularity of words in a set of apparently meaningless and chaotic sounds, linking these words to semantic meaning and understanding the grammatical regularities that link words with each other and give them different meanings. At a low level of complexity I can reproduce the narrative of the novel. At a higher level of complexity I can explain the plot, the characters and the dynamics between the characters. Even further up, I can extract an understanding of the deeper moral messages and life lessons that emerge from the book.

Machine learning follows similar mechanisms. It is a scaled-up version of existing statistical techniques of correlation, regression and Bayesian inference that enables recognition of patterns and structures in data and, under certain conditions, can detect causality in patterns. Machine learning is a label for a class of automated applications of these statistical techniques whereby many different structural models are tested[4] and continuously improved when new data come in. It tests its own findings and compares different interpretation models to select the best performing models.

This learning model explains an important economic characteristic of digital data: economies of scope. When two sets of data are partly overlapping, the cost of extracting knowledge from the two together is lower than the cost of doing this separately for each dataset. For example, studying the French language and French literature can better be done jointly than separately because there are large overlaps in the two knowledge sets. By contrast, studying French and Chinese jointly probably yields very little economies of scope because the two are not related. Economies of scope can also be applied to the benefit side. Studying economics and chemistry jointly will probably not yield more insights jointly compared to learning them separately. By contrast, studying micro- and macro-economics jointly would yield more insights than doing so separately.   Economists refer to this characteristic as "economies of scope" (Rosen, 1983).

A more general formulation of economies of scope could be illustrated as follows. Consider a structured dataset that consist of rows (number of observations on variables) and columns (number of variables). If the dataset is split in two parts (row-wise and/or column-wise), applying statistical inference methods to the two parts separately will not produce the same insights and robustness of results as applying them to the joint dataset, provided that the two parts are to some extent correlated – i.e. there is a relationship between the two parts. Merging two datasets with complementary data may produce more insights than keeping them separate. Economies of scope do not apply to completely unrelated datasets.

For example, machines can discover patterns in very large datasets that are beyond the cognitive capacity of humans to handle, though the machines often need human support to discover these patterns. The algorithms learned from one dataset may in some cases be transposed to other datasets. Learning obtained in a smaller dataset can be extended to expanded versions of the

---

[4] Vigo (2013), Boisot & Canals (2004) and Deutsch (2011) argue that looking for structure or regularities in data is not enough to understand or acquire knowledge. Knowledge cannot be derived through induction alone; it requires a theory or a prior framework that can be tested.

dataset. Extension to adjacent data areas can also generate economies of scope. The phone data can be overlaid with maps and with shops & restaurants data; applying the same algorithms and building on the observed patterns in phone data can produce even more insightful patterns, on top of those already observed in the phone data. Combining it with pay data in shops & restaurants adds further insights, etc. Applying machine learning algorithms separately to each of these datasets may be more costly and would not produce the same complexity of insights.

Economies of scope generate cost savings in data collection and analysis. They explain data trade and data-driven mergers. More generally, they explain why data-driven firms are so data-hungry and collect all the data they can get. There are countervailing forces at work. Economies of scope do not go on forever and are subject to diminishing returns. Scattered empirical evidence suggests that in some cases diminishing returns may set in at a very early stage (Pilaszy & Tikk, 2009, on film selection) while in other cases it only arrives when the number of observations increases many orders of magnitude (Varian, 2014) or never (Lewis & Rao, 2015, on the efficiency of online advertising).

2.3. Interoperability and barriers to access

Until very recently, human societies stored information on many different material carriers and in many different formats. Text was stored on paper using an alphabetical information protocol with 26 characters in the Latin version. Pictures were stored on canvas paintings or on celluloid film, and temperature could only be read from the length of a mercury column in a narrow glass tube, not to speak of time that could only be registered with a very complex mechanism of moving parts. There was no easy way to transport these material carriers of information around the world. Transposition between two formats required the intervention of a human interface to interpret and translate the information. This implies that not only the cost of production but also the marginal cost of storage, use, distribution and transposition were high.

That changed with digital technology. The digital format can be stored, replicated and transmitted electronically, much faster and cheaper and at a much lower energy cost than any analogue information system. Instead of transporting the material carrier of analogue information (books, films, thermometers) it now suffices to transport the information content in electronic format in order to reconstruct the information in a humanly readable format at the other end.

Low cost and high speed long-distance transmission already became available a century earlier with the telegraph, telephone and radio. But these electronic communication systems missed an essential feature: a common binary format. Modern digital information reduces information to its most elementary expression: a binary format with two states only, 0 and 1 – the minimum number of discernible states required to detect information. Further reduction is not possible: a one-digit system cannot distinguish between two different states and therefore contains no information at all. Binary digital systems are a common and universal information format. This shared information format facilitates transposition of information and connectivity between different digital devices and convergence between information storage formats in different devices. As a result, digital datasets can in principle be easily connected; they are interoperable.

Palfrey & Gasser (2012) put the question of interoperability at the center of the information society debate. Connectivity creates an enormous potential for spreading and sharing information that could deliver many social and economic benefits. But it also generates many perils if not well managed. The wish to avoid interoperability and easy access to datasets explain why there are so many man-made technical and regulatory barriers to prevent access and seamless connectivity. Some barriers to connectivity are desirable, for instance to protect privacy or the commercial assets of companies. Others may have detrimental effects, for instance when privacy creates security risks or when the lack of interoperability reduces competition between service providers and lock users into a specific system. Palfrey & Gasser argue that government intervention and regulation is required to reduce interconnectivity and put barriers to access in areas where it is socially desirable. Regulators face the difficult task of searching for an optimal degree of interoperability that realises the promises and avoid the perils of connectivity.

The tension between unlimited connectivity and access in digital data, and private and social preferences to limit access and retain a degree of exclusive ownership, is at the root of the debate on data ownership that is discussed in the remaining chapters of this study.

3.   The legal framework for data ownership and access

3.1.   Legal ownership rights in data[5]

Material goods are, by nature, rival goods. If one person uses it, the other cannot use it at the same time. For example, a CD or DVD cannot be played on two players at the same time. Rivalry makes it easier to claim exclusive property rights on material goods: you have it or you don't. Ideas, data and information are non-rival. Many people can use the same data at the same time without any loss of information content for any of these parties. Even if I have it, it doesn't exclude you from having it too. Non-rivalry has become easier in the digital age. Analogue datasets are usually costly to copy and re-use for other purposes; data ownership was not a major issue in the analogue information age. The dramatic reduction in the cost of conversion, copying and transmission of digital content between carriers significantly reduces the natural excludability barriers conferred on information by its material carrier. Just like low-cost printing technology triggered demand for copyright protection of writers[6], low-cost digital information technology raised questions about data free-riding and the protection of ownership. Reduced excludability is at the root of many property rights problems in digital information technology, including piracy of copyright-protected digital media products, privacy issues in personal data and private ownership of data.

Ensuring excludability of non-rival goods requires technical and/or legal intervention to define and attribute exclusive property rights. There is a well-established legal framework for exclusive intellectual property rights to intangible and non-rival ideas, as in patents and copyright for example. The legal status of data ownership rights if less clear. Partial and limited ownership rights to data are defined in the EU Database Directive (1996) and the General Regulation for the Protection of Rights in personal data (2016), combined with some provisions in the Trade Secrets Protection Directive (2016) and in general contract law.

Heverly (2003) argues that in order to determine whether data or information is property that can be owned by a natural or legal person, we must examine to what extent the rights given to information are analogous to those in more established property rights settings. These include the right to use, to exclude access and to transfer. He sees information or data as subject to ownership regimes similar to copyright, patents, trademarks and trade secrets. The underlying issue in all these regimes is a dynamic relationship between limited exclusive private rights and exceptions for common use or access – a hybrid ownership regime labelled as semi-commons.   There are several schools of thought in law on the origin of these limited but exclusive private rights. "Single variable essentialism" argues that the right to exclude is enough. "Multiple variable essentialists" say that the right to exclude is a necessary but not a sufficient condition and more is needed, including the right to use and transfer. "Nominalism" presents a more pragmatic approach and views property as whatever the legal system decides to call property. Nominalism emphasizes de jure ownership

---

[5] For a more detailed discussion, see European Commission (2017a, pp 19-22). See also European Commission (2017b).

[6] The Statute of Anne, also known as the Copyright Act (1710).

rights, as defined by law. Essentialism emphasizes de facto ownership rights: I own it when I have control over it. This seems to be the case for much of data ownership in the digital economy. This is often the case for ordinary goods too. However, Drexl (2016) is cautious about extending the traditional concept of property in Civil Codes in many countries to data ownership because non-rival data have different economic properties compared to tangible physical goods: someone else can use the same data without harm to the use of the original data. Intellectual property rights were invented precisely to deal with this non-rivalry.

a)   Ownership rights for data producers: the Database Directive

In the EU, efforts to create exclusive ownership rights in electronic data started with the Database Directive (1996)[7]. It was justified as an incentive for EU firms to invest more in the production of electronic databases and help the EU to catch up with other countries, in particular the US, in this respect. It takes a dual approach to data ownership. It gives full copyright protection to original databases that are the result of a creative human effort and a "sui generis" right for non-original databases, limited to 15 years. The Directive does not define the minimum human intellectual effort required to meet the standard of a creative database (Schneider, 1998; Hugenholtz, 2016). In any case, full copyright protection does not apply to the contents of the database ("the data") but only to its structure. The sui generis right protects the data as such but applies only if "there has been a qualitatively and/or quantitatively substantial investment in either the obtaining, verification or presentation of the contents" (Art 7(1) of the Database directive).

This dual approach reveals a fundamental problem in the use of copyright for the protection of data ownership. In the digital age, most datasets are generated by hardware and software, by machines, not by creative human effort. Data are representations of observed and measured facts, not creations of the human mind. The human creative effort is directed to the measurement methods and the design of the data machines, software and algorithms (covered by patents and copyright) but not their day-to day operations. The Database Directive (1996) contains a provision that extends protection in the EU to databases owned by non-EU firms, provided that similar protection is granted to EU databases in the country of origin of the non-EU firm. This triggered attempts in the US to introduce a database ownership right that mirrors the EU database right. Several proposals were submitted to Congress, without success (Zhu et al., 2008). A major factor that contributed to this failure was the absence of a consensus on need for copyright-like ownership rights on data.

The European Court of Justice (ECJ, 2005) considerably reduced the scope of the right in the British Horseracing Board case. A substantial investment in a database is not a sufficient condition to claim protection under the Database right[8]. The investment has to refer to resources used to make databases and seek out existing materials, not to resources used for the creation of data. That would

---

[7] There were some early legal initiatives to protect analogue datasets, such as directories and catalogues. Limited forms of database rights exist in Germany (as part of competition policy) and Scandinavian countries (for catalogues). Database rights also exist in Japan and Korea (see OECD, 2016, p 24).

[8] Only Australia recognizes the "sweat of the brow" substantial effort in constructing a database as a sufficient condition for copyright protection. In the Feist vs Rural case, the US Supreme Court rejected copyright protection for databases that lack originality in the arrangement of the data, and thus rejected that investments in creating a database are a sufficient reason for copyright protection.

rule out protection of data collected through sensors – essentially all electronic digital data. The right only protects the database investor against substantial harm to his investment. It protects the database as a whole and not specific data in the set. Extraction or re-use of data that does not harm the original investment is allowed. The ECJ thereby prevented a wide interpretation of the database right that would risk protecting the simple collection of information and facts. Drexl (2016) concludes that "it is quite obvious that the Database Directive is based on database technology that no longer corresponds to the use of data in an era of the Internet of Things, in particular … (15 years protection) … is much too static to adequately respond to the features of … real-time data services".

This ECJ interpretation has important implications for the economic interpretation of a database ownership right. It opens the door for a debate on the extent of substitution between upstream data production and downstream uses (see Section 4.1). Without substitution there is no risk for harm to the original data owner. It shifts the debate away from protection of the right to remuneration of the owner/creator ("*droit d'auteur*") that marks copyright protection, at least under the civil law tradition in Continental Europe, and towards a more economic interpretation that focuses on the protection of innovation that is more in line with the common law interpretation of copyright as well as with its economic interpretation. The economic purpose of copyright, or any intellectual property right for that matter, is not to maximize the revenue of rights holders but to achieve a socially optimal level of innovation. The two are not necessarily correlated.

Most importantly, it questions the need for data ownership rights as an incentive for the production of data. There are several reasons why the incentive argument for data ownership remains weak. First, Hugenholtz (2003) already argued that data are often a by-product of profitable economic activities and do not require additional incentives. Many online business models revolve around data collection as a by-product of information exchanges and commercial transactions. The commercial value of the data is monetized through services sales and ad auctions, no data sales. A car manufacturer has incentives to collect data from sensors in cars, irrespectively of selling the data. Second, the absence of direct data sales implies that data can more easily be made excludable through technical protection. Unlike in copyright-protected media products that are meant to be distributed widely, data are not necessarily widely shared and may not need legal protection to make them excludable. Third, the value of many data resides in their immediacy and real-time availability. It may depreciate very rapidly in the modern data economy. Extending data protection over many years may have little additional value. Data users require constant interaction with the provider and are unlikely to risk free-riding on that relationship. Finally, the database directive protects "data" but not the valuable semantic content, information and insights that can be extracted from the data. Content is extracted by means of data analytics which is usually the most costly but also the important value-added segment in the data value chain.

The weakness of the incentive argument that was a key element in the justification of the Database Directive would explain why data ownership rights are not so relevant for many data economy firms. The European Commission's (2005) own evaluation of the Database Directive finds no evidence that

it had any significant impact on investment in databases in the EU[9]. The EU keeps running behind the US in terms of number of databases. From an economic policy perspective the maximization of social welfare from data requires maintaining a balance between data ownership protection and access rights. When the importance of investment incentives for data owners diminishes, access rights can be widened. These economic implications are discussed in Chapter 4.

Apart from legal instruments like the Database Directive, secrecy and technical access restrictions can ensure data excludability when the law fails or is deemed insufficient. However, these measures may fail. In that case the Trade Secrets Protection Directive (2016) offers protection against the unlawful acquisition, use and disclosure of trade secrets. This applies only to information not "generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question" (Art 2.1(a)). It does not apply to data shared with other parties or publicly available, for instance on web pages or in publicly accessible catalogues. It is not clear how this would apply to data collected by sensors in machines. According to Drexl (2016, p 22), it is already technologically outdated as far as the modern data industry is concerned.

Most data are currently traded under bilateral contracts. This may work well in many instances. Drexl (2016) recalls that markets for sports broadcasting rights also work with bilateral contracts and without specific ownership protection in most jurisdictions. The organizers of sports events ensure excludability because they control access to the events and can charge a price for physical access, including for broadcasters of the event. Television broadcasting can be limited to specific territories. This becomes more difficult with internet broadcasting however where loopholes may occur. Similarly, contract law leaves loopholes in data ownership. A data owner can sign a contract with a data user that forbids any distribution to or re-use by third parties. However, that contract is not enforceable towards third parties who are not signatories to the contract. In other words, once the data are out in the open, the data owner has no legal means enforce his rights.

b)   Rights for data subjects: the General Data Protection Regulation

While the Database and Trade Secrets Directives protect data collectors, usually firms, data protection legislation aims to protect individuals (the "data subject"). In the law & economics literature, there is a considerable debate on the merits of granting full ownership rights to natural persons to their personal data. That debate was partly generated by the EU Data Protection Directive (1995), the predecessor to the General Data Protection Regulation (2016). Economists see property rights as a means to reduce transaction costs and increase the efficiency of exchange. Cato & Mayer-Schonberger (2012) claim that the "notice and consent" rights under the EU Data Protection Directive (1995) are not very effective because the transaction costs of verification and enforcement are high compared to the value of the rights. The right to be forgotten may be more effective because the subjective value of that right can be very high in exceptional circumstances. Laudon (1997) argues that legal protection of privacy is outdated and a system based on property rights over personal information would better satisfy the interests of both consumers and firms. In

---

[9] The evaluation argues that a repeal of the Database Directive would meet resistance, in particular from the publishing industry, and may therefore be difficult.

contrast, Acquisti et al. (2015, p 453) argue that while the assignment of property rights is generally welfare enhancing, granting consumers the right to sell their personal data may undermine consumer welfare. Sellers in a monopolistic market try to improve their capacity to price-discriminate collecting personal information on consumers. Marginal consumers in a monopolistically priced market make no surplus on their consumption and will be willing to sell their personal information for any marginal price. That enables the seller to collect more data and improve price discrimination. The market unravels as all consumers gradually move to the marginal position with improved price discrimination and will reveal their preferences. The monopolistic seller ends up with perfect price discrimination information across the entire market and acquires all consumer surplus. A counter-argument occurs when transaction costs are brought into the picture. In that case consumers would not sell their data unless the price exceeds transaction costs. In practice, we observe many consumers exchanging their personal data on websites in return for obtaining information. This implies that consumers perceive this exchange as producing a positive consumer surplus.

The new EU General Data Protection Regulation (GDPR, 2016) chose to preserve the European data protection acquis also embodied in the Council of Europe (1981) Convention on the Protection of Individuals with regard to Automatic Processing of Personal Data[10]. It builds on the Data Protection Directive (1995) and provides a comprehensive regime for the processing of personal data. The GDPR deliberately does not consider full and transferable private ownership rights for personal data. It justifies the absence of tradable ownership rights in personal data on the basis of human rights arguments: privacy is a basic human right that cannot be alienated. It creates inalienable and non-tradable specific rights for natural persons including (a) the prohibition of data processing without a legal basis (e.g. "informed consent"), (b) the prohibition to use personal data for other purposes that those for which they were originally collected, (c) the right for the data subject to access and extract ("port") his personal data, and (d) the right to be forgotten. These rights are assigned to the data subject as a natural person and reduce whatever rights the data collector has as a creator of a database of personal data. These rights cannot be claimed by firms or legal entities whose data are collected by other parties (in a B2B context for example).

Art 4(1) of the GDPR defines personal data as "any information relating to an identified or identifiable natural person ('data subject')". The definition of personal data implies that there is a clearly distinguishable category of non-personal data. The rapid evolution of data collection and analysis technology may create ambiguous borderline cases in the definition of personal data. For example, non-personal data like temperature readings and electricity consumption in a house can often be linked to personal behaviour too. In social media platforms some personal data are intertwined with data generated by the platform and other persons operating on the platform. For example, social links and online review scores cannot be ascribed to a single person. They are the product of social interaction between many persons, enabled by the platform. The recently proposed EU Digital Content Directive (Com(2015)634) would allow consumers to "retrieve all content provided by the consumer and any other data produced or generated through the consumer's use of the digital content" (Art 13 (2) c). This potentially expands the provisions of

---

[10] Council of Europe (1981) European Treaties series nr 108.

personal data portability in the GDPR from data provided by the individual to data generated through the consumer's use of a platform including metadata generated by the platform[11]. These ambiguities in the scope of the definition of personal data, combined with the absence of personal data ownership rights, may create uncertainty and increase transaction costs in personal data exchange situations.

The GDPR restricts aggregation and trade in personal data and thereby puts constraints on the potential benefits from economies of scope in data aggregation. Art 5(b) states that data can be collected for specific, explicit and legitimate purposes but should not be processed for any other "incompatible" purposes. A secondary purpose is not prohibited but should not be "incompatible" with the original purpose. Art 6(4) lists a number of conditions for re-purposing. The European Data Protection Supervisor has tried to clarify these conditions. For example, incompatibility may occur if two uses of personal data (within a single firm or between two firms) are in separate and non-substitutable markets (EDPS, 2014, p 27). The proposed Digital Content Directive (COM(2015)634) complements the GDPR in this respect because it recognises that consumers share personal data in return for access to digital content (Art 2), provided this is done in full respect of the GDPR. That provision gives legal recognition to one of the most widespread data trade business models in the digital economy. The GDPR does not apply to anonymized datasets. However, many anonymized datasets are vulnerable to de-anonymization attacks[12].

There are differences between the legal and economic interpretation of the GDPR. From a legal perspective, the GDPR gives data subjects no full ownership rights, only certain specific rights including the right not to be subject to data processing without a legal basis (e.g. "informed consent"), access, limited re-purposing, the right to be forgotten and the right to data portability. The data collector ("data controller" in the terms of the GDPR) has a fiduciary role and should ensure the respect of the specific rights of the data subject under the GDPR. The granting of specific rights to data subject implies that any remaining residual rights not included in the specific rights in the GDPR accrue to the data controller. In the economic literature on property rights, residual rights are defined as the rights that remain unspecific after specific rights have been assigned to other parties. These residual rights are called property rights. (Grossman & Hart, 1986), From an economic perspective the assignment of residual rights to a party called "the owner" is a way to reduce the cost of contracting, since contracts (or laws in this case) are by nature incomplete, or too costly to complete. In that sense, the GDPR de facto (but not de jure) assigns property rights on personal data to the data collector, however limited they may be due to his fiduciary role. In reality, data subjects exchange their personal data in online markets, for example when they access "free" online services in return for letting the service provider or data controller collect some personal data. In these cases the data subject retains the specific rights on his data as defined in the GDPR; the service provider acquires the residual rights.

---

[11] This Directive is currently debated between the European Commission, the Parliament and the Council and no final decision has been taken yet. In particular, the proposed Art 2 is contentious and may not make it to the final reading.

[12] See Ji et al. (2016) for a survey of anonymization and de-anonymization techniques.

<u>c)   From the legal to the economic debate</u>

Clearly, ownership rights in data are only very partially defined. The Database Directive gives some limited property rights to data collectors, inspired by copyright but limited in scope by ECJ jurisprudence. The GDPR gives some specific rights to data subjects but refrains from defining a residual ownership right in personal data. This leaves a wide area where ownership or residual rights are not legally specified, or incompletely specified. Exclusive data ownership thereby becomes a de facto right: I have the data and can effectively prevent others from accessing the data, therefore I am the owner of all residual rights not explicitly assigned away to other parties through specific legal or contractual rights. Data collection and protection technology, combined with the market power of data firms and the willingness of data subjects to incur (opportunity) costs to protect or trade their personal data, become the main drivers of data exchanges and market outcomes in data markets.

The key economic policy question is whether this incomplete legal regime, with a combination of limited database rights, the protection of trade secrets and non-transferable rights to personal data protection, is sufficient for the needs of the digital economy?

An active legal debate has emerged, mainly in Germany, on the merits of filling the gaps in data property rights and the creation of a data ownership right. According to Drexl (2016) the debate is especially alive in Germany where there is considerable discussion on the impact of the Internet of Things and the increasing use of digital data in industrial applications. This is often referred to under the label of "Industry 4.0" or the "Fourth Industrial Revolution".[13] Moreover, Germany has a strong car manufacturing industry where data ownership issues are coming to the forefront with the digitization of many driving and traffic management services. In line with Hughenholtz (2003), Kerber (2016) observes that there is no evidence of an incentive problem regarding the production and analysis of non-personal data. In particular, he argues that while data are non-rival in use they are characterized by excludability as it is technologically feasible to keep them secret and protect them against copying and leaking to the public. He argues that private contracts are direct solutions to the danger that trading or granting others access to data pose to the excludability of data. Kerber (2016) concludes that data ownership rights are not recommendable on the basis of innovation and IP economics arguments. Socially optimal incentives to produce and analyze data, and the adequacy of contractual obligations and technical restrictions to ensure the excludability of data, are empirical questions in his view. He therefore calls for further evidence-based research on these topics that are surprisingly understudied in the economics literature. In line with Kerber (2016), Drexl (2016) argues that the creation of a new system of data ownership is not advisable as factual control of data supported by technical protection measures allow data holders to exclude third parties or charge them a price for making data available. He also argues that there is no economic justification for the creation of a new system of data ownership based on the incentive argument.

Other authors accept the rationale for the creation of a data ownership right and discuss the practical legal modalities. Wiebe (2016b) suggests that a data ownership right could be a coding right (first storage or recording of data) that requires novelty and registration and should protect against copying but not against independent creation, for a maximum duration of 5 years. He points

---

[13] For instance, see Drexl et al. (2016a&b), Surblyte (2016) and Wiebe (2016a).

out that there is an allocation problem however: who gets the right? In this regard, he observes that it is typically difficult to identify the agent who makes the investment in the creation and analysis of data. It is often also difficult to assess who is the most efficient user of data. In addition, Wiebe (2016b) points to a specification problem: does a data ownership right lead to an unqualified indirect protection of information? It is in this respect that he acknowledges severe theoretical and practical impediments to a data ownership rights. Zech (2016) also addresses the incentive argument for generating and trading data. In line with Kerber (2016), Zech observes that the incentive function may be less important for scenarios where data can be produced at virtually zero marginal cost. However, he also acknowledges that in other scenarios huge ex ante investments in data production and measuring devices are required. This could lead to a market failure and make legal protection at the stage of data production necessary. In addition, Zech proposes that transferable ownership rights are assigned to "the economically responsible operator" of the measurement equipment. The author justifies the proposal with economic arguments related to data investment incentives and lower transaction costs in ex-ante contracting and ex-post risks. The "economically responsible operator" refers to Art 950 in the German Civil Code that creates an ipso iure ownership title on the results of a process of transformation of physical object(s) into a new distinct object, on condition that the value added by the transformation is not significantly inferior to the value of the object put into the transformation process. For example, cloth can be turned into a dress. It may be a bit hard to translate this into an economically meaningful interpretation on ownership rights. Am I the responsible operator of my smartphone or is it the telecom operator or the provider of the operating system or the provider of the app that I use? Who contributes most of the value to my smartphone data? Zech admits that this is an empirical question that can depend on the cost of data production as well as the different types of data (e.g. secret data, factually exclusive data without secrecy, personal data, structured arrangements of data, i.e. databases) and therefore cannot be answered by theoretical reasoning alone. As such, Zech's (2016) and Wiebe's (2016b) proposals only shift the question from "should we have a data ownership right" to "who will get that right". Both authors propose some legal characteristics for a data ownership right but do not clarify who should get it and why. Zech argues that ownership should only protect against commercial infringements, not against non-commercial re-use of the data. That puts his proposal in line with the ECJ judgment in the BHB vs Hill case. He does not distinguish between personal and non-personal data, which may lead to a conflict with the EU GDPR.

3.2. Legal aspects of competition in data markets

Besides the legal debate on the merits of granting exclusive ownership rights to data, there is also a debate on the reverse problem, granting access to exclusively owned data. Full ownership grants an exclusive monopoly right on the use of data. To the extent that the data are unique this could lead to distortions in the data market. The owner could decide not to sell his data at all, thereby blocking access to potential downstream users of the data, or extract a high rent from downstream users that could completely erode their surplus. This hold-up problem raises the question of competition in data markets and access to data.

Opinions are divided as to whether data distort competition between firms and would require intervention by competition authorities or regulators[14]. Some argue that data markets are active and offer many substitutable sources of data. Data do not constitute an entrenched source of advantage for firms because they are non-rival, ubiquitous, potentially excludable but highly substitutable and have a very short economic lifespan. They can be produced and distributed at near-zero marginal cost. There is a thriving data market with big players that amass massive volumes & variety of information (not only Google, Facebook, Amazon, Apple etc. but also pure data companies like Acxiom, Datalogix and Bluekai) as well as a myriad of smaller players that focus on niche markets (for instance most apps collect a wide range of personal data). Many of these data are actively traded and accessible and there are often several substitutes. They conclude that data is neither inimitable, nor rare and non-substitutable. The short history of the digital economy has so far shown that substitutes exist. Competitive advantage is not acquired by accumulating lots of data but rather by developing the organisational capabilities to make better use of data.

However, it is also important to note that, although some data are not traded, this has not prevented competitors from entering downstream services markets for these data. Skype and Facebook are sitting on piles of social network data but Whatsapp, Snapchat and Instagram still managed to build up a powerful market position ex-nihilo and without having these data to start with. Thousands of taxi companies had valuable but totally fragmented data. Collaborative economy platforms like Lyft and Uber designed new taxi services that managed to construct aggregate data and moreover added trust and reputation information to it.

Others argue that many data are not traded. There may be alternative sources for consumer data but less so for industrial (non-personal) data. Data-driven mergers are rapidly increasing, not only driven by cost efficiency but also by economies of scope and scale. Mergers can stop competitors from snooping up the market. Preventing data portability and interoperability creates barriers to entry and limits competition. Firms, in particular multi-sided platform markets, can use data to strengthen their competitiveness in a market. For example, Booking.com and Expedia aggregate hotel booking data across many hotels and cities. That brings more transparency and competition in the hotel market and reduces transaction costs for platform users. It may also give platforms leverage over hotel price setting, extract higher margins and boost their own revenue. However, many hotel reservation sites collect similar data and compete in the market. The success of a website is not only due to data; commercial strategies and service quality matter as well. Compelling firms to share the source of their advantage may lessen the incentive to invest in those facilities.

The legal options for redress under competition law against the refusal to trade unique and non-substitutable data owned by a firm in a dominant market position are not great according to jurisprudence. A number of US cases (LiveUniverse vs MySpace, Trinko, Facebook vs Power Ventures) show that US courts insist on a pre-existing voluntary course of dealing and proof that the monopolist is willing to forgo profits in order to achieve an anti-competitive end. In the EU, Article 102 of the Treaty seems to offer more scope when access seekers needs data as an input for a new

---

[14] See, for instance, Lambrecht & Tucker (2015), Tucker & Wellford (2014), Graef et al (2015a), Stucke & Grunes (2016), Strucke & Ezrachi, 2015, Sokol & Comersford (2016), Ohlhausen & Okuliar (2015) and Evans & Schmalensee, 2014. For a more detailed discussion of these authors, see Martens (2016), chapter 3.

product that does not directly compete with the main product produced by the data owner. The Magill, IMS Health and Microsoft cases provide some jurisprudence in that direction. The ECJ sets out four conditions for regulatory action based on competition law principles[15]: that the data is indispensable for the downstream product, that there would not be any effective competition between the upstream and downstream product, that refusal prevents the emergence of the second product, and there is no objective reason for the refusal. Indispensability remains hard to prove in a world of ubiquitous and substitutable data. There are as of today no competition cases in the EU or in the US related to (the absence of) trade or market positions in data. Competition cases that could have touched upon this issue, such as the Facebook-Whatsapp and DoubleClick-Google mergers, were careful to avoid the perception that data could constitute a competition problem. Competition authorities in the EU and in the US have so far not found any competition problems related to big data. Case law does not support the contention that data collection is an antitrust problem. The nature of the relationship between platform users and data collectors is more likely to fall within the realm of consumer protection law (including privacy and data protection law) than competition law. Online data have generated unprecedented consumer benefits in terms of free online services, improved quality of services and rapid innovation. The ability to offer free services via monetization of data sales and advertising is mostly seen as a pro-competitive effect and not harmful from a competition perspective. The absence of monetization would reduce the volume and increase the cost of online services and reduce competition in product markets.

Drexl (2016) extensively discusses the questions of whether remedies to promote data access are already provided by EU competition law and whether there is a need for legislation on data access from a competition law perspective. Drexl argues that while EU competition law provides some remedies to promote data access and to address excessive pricing it shows significant flaws with regard to the data economy. For instance, EU competition law only addresses a particular kind of market failure, i.e. identifiable anti-competitive conduct is banned *ex post*. In contrast, as Drexl (2016, p. 44) argues, competition law enforcers are typically not able to regulate markets *ex ante* by "imposing positive rules of conduct in form of behavioural remedies that require on-going monitoring". In addition, information asymmetries with respect to the value of data (and big data analytics) are a potential source for market failure that cannot be addressed by competition law. More specifically, market failure may occur as contractual negotiations about data access fail because the purchaser of data cannot properly assess its value.[16] In the light of these arguments Drexl (2016) concludes that EU competition law is unlikely to provide sufficient remedies to promote data access. However, while Drexl (2016)'s analysis calls for actions that go beyond competition law he also observes that competition law thinking is instrumental in developing additional pro-active and pro-competitive regimes to promote data access. For instance, data portability rights are likely to enhance competition when factual control of data leads to a lock-in effect.[17] Block exemption regulations provide another example of how to create competition by preventing anti-competitive business practices. Finally, Drexl (2016) argues that progressive sector-specific legislation in fields

---

[15] For instance, see Drexl (2016).

[16] See also Section 5 on data trade for a more detailed discussion of this phenomenon that is known as "information paradox" (Arrow, 1962).

[17] See also Section 6.2. on data portability and interoperability for a more detailed discussion on lock-in effects.

such as environmental, public health or road traffic law may be adequate to develop models for legislation on data access over time.

4.   The economics of data ownership

In this section we explore the policy questions (a) if the incompleteness of the current legal framework for data ownership leads to data market failures and (b) whether better-defined data ownership & access rights would improve market outcomes?

The European Commission Communication on the Data Economy (2016) focuses on non-personal machine-generated data. It points out that there is little legal protection of ownership for "machine-generated data". Ownership protection under the sui generis right in the Database Directive (1996) is restricted. Firms have to solve legal issues in data access and trade through bilateral contracts. The GDPR gives some specific rights to the subjects of personal data but stops short of an ownership right.   This incompleteness of the legal framework is not surprising given the fact that fast technological advances in digital data collection, storage distribution and analysis technologies are very recent. Also, an incomplete or even a totally absent legal ownership regime does not mean that there are no ownership rights. On the contrary, a legal vacuum is usually filled up by de facto ownership rights. This has been observed before in situations where a legal framework was absent. In his paper "Might makes right", Umbeck (1981) explains how private land property rights emerged during the California Gold Rush in the complete absence of any legal and state-enforced property rights. An equally distributed violence technology among miners (they all had guns) ensured that none could dominate the others and resulted in a fair distribution of mining parcels (Skaperdas, 1991 & 1992). As the title of the paper suggests, in the absence of the law the power distribution between parties will determine ownership rights. Ellickson's (1986) study of cattle ranchers and farmers in Shasta County shows that, even when legal rights are present, the actual outcome may differ substantially from these legal rights when power is asymmetrically distributed.   Power does not necessarily mean the threat of violence as in the California Gold Rush. It may consist of more sophisticated ways to re-distribute costs and benefits between parties. The absence of legally defined rights, or the presence of high enforcement costs for these rights, creates room for bargaining or strategic behavior between parties to settle the allocation of rights. In the absence of outright violence, the outcome depends on economic powers to appropriate benefits and inflict costs. Enforceable legal ownership rights will reduce the margins for bargaining; they will rarely eliminate that margin. Transaction and enforcement costs in legal rights will widen the margin.

In economics, property rights are defined as residual rights, or costs and benefits that are not specified in a contract; they are allocated to a party called "the owner". Specific rights, clearly identified in a contract, may be allocated to other parties. Contract theory in economics has two branches. Grossman & Hart (1986) assume that complete contracts that specify all possible outcomes can in theory be written but parties refrain from doing so because this would entail high transaction costs. Another branch (Laffont & Tirole, 1993) assumes that complete contracts are not possible because of inherently incomplete information. In both cases, contracts generate a degree of uncertainty about outcomes or unresolved residual claims that become the subject of bargaining. The advantage of having a legally specified ownership right is that the law allocates these residual rights to a party. In the absence of such a law, parties start bargaining over the allocation of these residual rights. Even in the presence of a law, uncertainties may exist and lead to bargaining in front

of a third-party judge who will settle the outcome. When transaction costs to go to court are high, parties may settle the dispute out-of-court "in the shadow of the law".

For example, the EU GDPR (2016) allocates only a limited number of specific personal data rights to data subjects. It leaves all other rights or residual ownership rights unspecified. That creates a bargaining situation between the data subject and the personal data collector. In the data economy this often results in an exchange deal whereby the collector provides the requested information to the data subject in return for being allowed to collect personal data about the subject, for instance through cookies in a web browser. From an economic perspective, the data collector becomes the de facto owner of the residual rights to personal data, even if the law does not specify or even rejects an ownership right. His rights are limited by the specific rights assigned to the data subject but he can claim the costs and benefits of everything not covered by these specific rights. While this is a fairly common and usually benign example, the bargaining situation can get more conflictual, for example in the case of "machine-generated data" where the data provider claims exclusive ownership on data collected by a machine owned by the data subject (See section 4.3).

Economists are inclined to think that well-defined and easily enforceable property rights are an efficient way to organize an economy. They reduce uncertainty and the margins for bargaining, thereby reducing transaction costs that create deadweight welfare losses for society. That, in turn, avoids market failures. While the legal debate gets stuck in the question to whom an ownership right on data should be given, economics sees easily transferable property rights as a sufficient condition for an efficient allocation of resources. According to the Coase Theorem (Coase, 1960), it does not matter which party receives the initial allocation of property rights. The rights will end up in the hands of the party that attaches the highest value to these resources, provided that transaction costs remain relatively low compared to the value of the right. High transaction costs can however block a welfare-enhancing transfer and lead to market failures. In other words, in order to answer the question if the current data property rights regime – including all the gaps in this regime – is sufficient, economists shift the question from "who will get the right" (Zech, 2016; Kerber, 2016) to "are there potential market failures that prevent an efficient allocation of that right"? If such market failures occur the question is whether they can be addressed through legal and regulatory intervention that re-defines or re-allocates property rights.

In this chapter we look at two potential sources of data market failures. We start in Sections 4.1 and 4.2 with the basic economics of IPR and a situation where data ownership rights are well-defined. However, there may be externalities that are a typical source of market failure. The treatment of this market failure fits well with economics of intellectual property rights models that can be applied to data. It also fits well with the copyright-like protection for data owners under the sui generis right in the EU Database Directive. In Section 4.3 we move to strategic behavior and bargaining between data rights holders as a source of market failure. This may occur when ownership rights are well-defined but split between two parties – the typical anti-commons situation. It may also occur when rights are poorly defined and parties start bickering over these rights. Irrespective of the cause of bargaining, it often leads to inefficient outcomes. We borrow concepts and tools from game theory to explore outcomes. High transaction costs are a third source of market failure that interacts with these two sources.

4.1. Applying the economics of intellectual property rights

The previous chapter explained the uncomfortable legal relationship between data ownership rights and copyright because, for historical reasons, copyright law is linked to human creative work. Until recently humans were the only source of creative activity and the extraction of knowledge from less structured information sources. With the arrival of digital technology and artificial intelligence humans no longer have a monopoly on information processing and the extraction of knowledge from information. From an economic perspective however we are not bound by these historical legal interpretations. We can explore to what extent IPR-like mechanisms are suitable tools to overcome the problems of non-rivalry and market failure in the supply and trade of data. In other words, are the economic characteristics of data markets comparable to, or different from, markets for intangible intellectual property? This section discusses the pros and cons of such an approach.

The assumption that underpins the IPR approach is that data production has a cost and requires a financial incentive in order to stimulate investment in data collection, storage and analytics. Data are non-rival and non-excludable by their very nature: many parties can use them at the same time without any loss of utility for any of the parties. If they are not made excludable by law they become a public good. Free-riding by other users would take away incentives for private agents to invest in their production. That would result in a market failure because of undersupply of socially useful data. In Section 3.1 we explored a number of reasons why these incentive assumptions may be unrealistic in many cases. Data are often a by-product of commercially profitable exchanges and/or they are protected by technical measures and not revealed to trading partners. Even when they are traded directly, their value often resides in real-time supply that involves repeated interaction and leaves little room for free riding. In this Section we stick to these assumptions however and explore the underlying mechanism. The main purpose is to show the balance between static welfare losses and dynamic welfare gains in this IPR-type model. It allows us to conclude that weak incentive effects can translate into wider access rights to data without social welfare losses. In the next Section we relax these assumptions and introduce bargaining.

The standard economic model of intellectual property rights revolves around the trade-off between static social welfare losses from over-protection of ownership and dynamic welfare gains achieved through the incentive effect for more investment in production of innovative content (Besen & Raskind, 1991; Landes & Posner, 2003; Posner, 2005). The trade-off is presented in Figure 3. An ownership right to a dataset gives the owner a monopoly on decision-making regarding the use of that right. He will fix a price to sell or license the use rights to others. The white rectangle represents the revenue for the data owner. The yellow triangle on top is the consumer surplus derived from the data, or the difference between the price paid by users and the actual value of the data for them. That price will maximize the owner's profit but not the aggregate welfare of society. There is a deadweight welfare loss associated with this monopoly (the triangle to the bottom right). If the owner would be able to differentiate prices according to the ability and willingness of his customers to pay for the data, the deadweight loss could disappear. It would also shift all value from consumers to producers. In practice, price discrimination is often difficult and only applied within a limited range. The basic economic premise behind IP rights is that the short-term welfare loss is compensated by long run gains generated by a continuous stream of new innovations or, in

this case, new datasets that will be produced because of the financial incentive to invest in more data collection.

The task of policy makers is to minimize the deadweight losses to society. They can do so by modifying the scope of the data ownership right and strike an appropriate balance between short-term losses and longer-term gains in further innovation. Modifying the scope affects the price and revenue for the owner and thus deadweight losses for society. An important implication of this economic mechanism is that higher (lower) protection is not necessarily better (worse) for society. There is an optimal degree of protection (Figure 4). Higher protection may benefit the rights holder but may have a negative impact on balance between static welfare losses and dynamic gains. Excessive protection would reduce downstream innovation and the production of complementary products. Higher protection is warranted when the production of data requires a stronger financial incentive because of high costs. As explained above, this may not always be the case for data, quite to the contrary. That changes the balance in the equation between static welfare losses and dynamic welfare gains and would lead to the conclusion that less protection is required in order to maximize benefits or, in other words, granting wider access to data would not be harmful to data production. This leads the OECD (2015) to argue in favour of open data that are widely accessible by everybody, without restrictions that may cause obstacles to access.

Figure 3: The static economic welfare effects of IPR
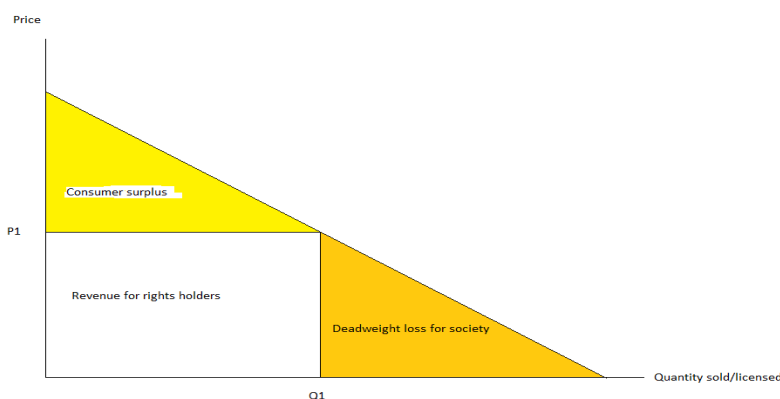
Open access does not necessarily mean free access at a zero price[18]. If perfect price discrimination is feasible wide access is ensured because all users can pay their use value to the data owner, provided that there is no possibility for arbitrage and re-sale or re-use between lower and higher-paying users. Preventing arbitrage requires a combination of legal, contractual and technical protection measures. It becomes more difficult to avoid welfare losses when price discrimination is difficult and

---

[18] See OECD (2016). Maximizing the economic and social value of data, p. 15) suggests that "data commons is often misunderstood, and used as a synonym for "open data", the distinction between data commons and open data is a small but significant one."

arbitrage is easy. In an extreme case where excludability cannot be ensured after a transaction, it may better be for the data to be owned and used by a single party that attaches the highest value to it.

Figure 4: The diverging interests of rights holders and society
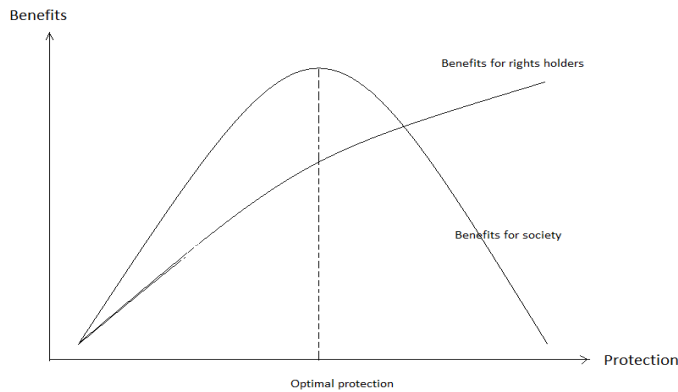
Patent protection is a good example of this balance. Technological innovation is often a cumulative process where new inventions build on and combine technologies developed in previous inventions (Scotchmer, 1991; Arthur, 2008). Limitations to the scope of excludability facilitate downstream innovation that builds on the invention protected by the patent, despite the fact that these limitations may deprive rights holders of a substantial stream of revenue when the patent expires. This explains why patent protection is very limited in time. Rights holders may actually lose a substantial amount of revenue when the patent expires, for example on popular pharmaceuticals. Still, it is beneficial for society to let the patent expire because that enables complementary and competing innovations to come onto the market. Maintaining patent protection for too long would slow down downstream innovation. This would upset the balance between static losses and dynamic welfare gains and allocate too much of the benefits to rights holders, at the expense of welfare for society at large. By contrast, the production of copyright-protected media products like music, films and books is only weakly cumulative[19].

---

[19] Some limitations to the scope of copyright protection may enable more cumulative innovation. For example, digital technology has facilitated text and data mining of online databases, newspaper archives, store catalogues and social media websites, and scientific publishers' databases. The mined data are used for research purposes and can contribute to innovation. The EU copyright regime provides an exception to copyright protection in case of text and data mining for non-commercial scientific purposes only.

4.2. Data as intermediate goods

A similar cumulative or downstream innovation argument can be applied to data. Koutroumpis et al. (2016b) distinguish between intermediary and final data goods. Media content and software for example are mostly final consumer goods. They are heterogeneous products that compete with each other and boost welfare-enhancing variety for consumers. Because they are competing substitute products that vie for a share in the same market, and because media production and innovation is only weakly cumulative, strong copyright protection for media products and software will not block downstream innovation. Data however are mostly intermediary goods that are used in production processes by other parties. There may be few substitutes available and the production process can be strongly cumulative. Strong and long-term protection may disable downstream innovation in new complementary products and services. For this reason the EU Database Directive accords only 15 years of protection to data under the sui generis right. However, 15 years may still be close to infinity for some types of data in the digital information era.

The degree of substitution between upstream and downstream data applications is a key issue in setting an appropriate degree of protection for databases. If the re-user uses the original data to produce another database that is functionally different and therefore a complement rather than a substitute to the original, the upstream and downstream users will not compete in the same market and revenue for the original creator of the database will not be affected by downstream uses. However, to the extent that the databases are functionally (partially) overlapping, there may be a revenue loss for the original creator that may negatively affect the incentive to invest.

Zhu et al. (2008) present an economic model that explains the conditions that should be attached to ownership and re-use of a database. They examine the history of database legislation and jurisprudence in the EU and US and extract three factors that have played an important role: substantial expenditure for the creation of an original database, the extent of functional equivalence between the reused data and the original, and injury for the original creator. These can be translated into economic concepts: fixed investment costs for the creator, substitutability versus complementarity between the original and reused data, and impact on the revenue of the creator and re-user. They bring these variables together in an economic model of competition and calculate the welfare impact (combined revenue of creator and re-user) of different access conditions for the re-user to the creator's data. The creator pays a fixed investment cost to produce the database. He sells the data at a monopoly price and maximises his profit. If the re-user uses the data to produce another database that is functionally different and therefore a complement rather than a substitute to the original, the two will not compete in the same market and revenue for the creator will not be affected. Consumers of both databases will benefit because they have more variety of data available. However, to the extent that the databases are functionally (partially) overlapping, there may be a revenue loss for the original creator. Consumers may still have a preference for one or the other database because the various aspects of the databases may only partially overlap. A market failure occurs when the revenue loss for the creator reduces his profits to zero. That will stop production of the original database. To prevent this, re-use should thus be allowed only up to the point where substitution reduces profits of the creator to zero.

A special case occurs when the original database is actually a by-product of another production process, with zero additional production costs for the creator (Hugenholtz, 2003). For example, eBay's second-hand market price data are a by-product of its auction activities. There is no risk that the original data collector will stop the data production process if he makes no profits on licensing access to that database. The database creator may want to maximize revenue from this data by-product and set a profit-maximizing price. That generates short-term deadweight losses and a long-term reduction in downstream innovation for society. To prevent this market failure, compulsory licensing or even unlicensed use exceptions could be added to the database ownership right. This feature was included in the original draft text of the EU Database Directive but subsequently removed. The US proposal HR3261 for a database right also included a compulsory licensing provision.

Zhu et al. (2008) conclude that the "substantial expenditure" requirement in the EU Database Directive should not be interpreted as an absolute value but as a relative value of the investment costs compared to the market value of the data. No protection should be given when the database represents a trivial expenditure (or simply a by-product of another production process) or when the re-user database is highly differentiated from and a complement to the original. Conversely, re-use should be forbidden when it is a close substitute and risks undermining the original database. These conclusions are in line with ECJ (2005) "BHB vs Hill".

4.3. Data ownership fragmentation and strategic behavior

Buchanan & Yoon (2000) and Bertichini et al. (2008) argue that ownership characteristics vary on a continuum from commons, to private property and anti-commons. Exclusive ownership rights turn non-rival and fully interoperable digital data from common property into private property (Fennell, 2011). That gives financial incentives to the data producer and owner. Privatization, however, may go too far and create a new problem of anti-commons or excessive fragmentation of ownership rights and reduced interoperability. This is especially problematic in the case of data where economies of scope in the aggregation of datasets cannot be realized when ownership is fragmented. Each owner holds exclusive rights to exclude other owners from realizing the potential benefits of data aggregation. Decisions by one owner affect other owners. This leads to strategic behaviour whereby owners try to internalise benefits for themselves and externalize costs to others. The uncoordinated exercise of exclusion rights leads to under-utilization of data.

A distinction can be made between horizontal and vertical anti-commons problems. Horizontal anti-commons problems occur when two owners hold complementary data that have more joint value than the sum of separate values because of economies of scope in data. While they know that merging the datasets would increase the joint value, they will haggle over the allocation of benefits. The exclusion rights are exercised simultaneously and independently by the owners. Schultz et al. (2002)[20] show that the Nash equilibrium[21] that is achieved under the two-owner scenario does not

---

[20] Their paper does not specifically focus on data ownership but the same principles and mechanisms apply.

lead to socially optimal value maximisation, compared to a situation where both datasets would be owned by a single agent. Vertical anti-commons occur when a downstream user depends on the supply of an upstream monopolistic supplier of the data. This case is explained in Zhu et al. (2008). If the downstream user produces a product that competes with and substitutes for the services of the upstream data supplier, the latter will have an incentive to threaten to cut the data supply, unless he receives sufficient compensation for potential revenue losses. Cutting the data supply would result in a (potential) revenue loss for the upstream supplier. He has an incentive to find a solution. This starts a bargaining process. The downstream service operator will be willing to pay a fee but a very high license fee may undermine his business model. In this scenario, both sides may find it individually optimal to only partly meet the demands of the other side. This may result in an overall suboptimal outcome that reduces the incentives of the downstream supplier to maximize investment in his service (Schultz et al., 2002).

If the downstream service is a complement to the service of the original data provider, the latter will want to maximize his share in the revenue of that service. This will again trigger a bargaining situation and a suboptimal outcome. The bargaining may result in a recurrent transfer payment with separate ownership or in complete vertical integration of ownership of the upstream and downstream services. For example, the Twitter Firehose service allowed downstream users to access the Twitter data and design new services around this. Twitter then cut the Firehose and transferred this data service to a new company that charged higher prices for data access. Hybrid cases can also occur. For example, the take-over of the Whatsapp and Instagram social networks by Facebook has elements of complementarity in the datasets as well as competition between services.

These situations can also be described in terms of an Ultimatum Game.[22] The upstream data producer makes a take-it-or-leave-it offer to the downstream data service provider. The offer splits the revenue from the downstream service in two parts, for the upstream and downstream parties. If the downstream party rejects the offer, neither party will earn anything. If he accepts the offer, both will have some revenue. In principle, the upstream producer could make any offer that still leaves some minimally positive revenue for the downstream service. It would be economically rational for the latter to accept. However, he may not find this a fair distribution or a worthwhile remuneration for his entrepreneurial efforts and decline the offer, leaving both parties empty-handed. The outcome of the game is unlikely to leave all revenue in the hands of the downstream party who will consequently reduce his efforts to invest in the development of the downstream

---

[21] The Nash equilibrium is a fundamental concept in game theory and has its origins in Nash (1950 & 1951). It is a solution concept for non-cooperative games, i.e. strategic interactions between players in which cooperative behaviour cannot be strictly enforced, e.g. via an explicit contract. To illustrate, in a Nash equilibrium, I am making the best decision I can, taking into account the decision of my opponent while my opponent's decision remains unchanged, and my opponent is making the best decision he/she can, taking into account my decision while my decision remains unchanged. In game theoretic parlance, a Nash equilibrium is achieved when each player chooses his/her best strategy given that the opponent chooses his/her best strategy and thus no player has an incentive to unilaterally deviate from this outcome.

[22] In the traditional ultimatum game two players try to agree how to split an amount of money between them. The rule of the game is very simple: player A proposes a split and player B can either accept or reject it. If B rejects it, none of the players get any money. If B accepts it, both players get their part in the proposed split. In theory, any split proposed by A would make B better off than getting nothing at all. Empirical tests show that splits whereby B gets less than half are usually rejected. This has often been interpreted as proof of irrational behaviour.

service which may lead to a suboptimal outcome where socially desirable innovative projects are not undertaken.

Schultz et al. (2002) do not discuss the impact of ownership fragmentation on the welfare of consumers. There are many cases where the fragmentation of data ownership between firms affects consumers. That impact is often ambiguous though. For example, fragmentation in ownership of consumer data affects firms that use consumer data for marketing and advertising purposes. Fragmentation benefits consumers when it protects their privacy. It harms them when it results in lower-quality services and less consumer choice. Ads can be very distractive and cause irritation, or they can be designed to persuade rather than to inform consumers and drive them away from their autonomous choices. Another example is driver data. Keeping my driving data away from my car insurance company may avoid higher premiums if I am a high-risk driver but may also prevent me from benefitting from a lower premium when I am a low-risk driver. On the other hand, integration of different sources of driver data for the purpose of traffic management would have an unambiguously positive impact on consumers.

A bargaining game may also occur between sellers who collect consumer data for marketing purposes and consumers who want to hide their personal information. Consumers get an information offer from sellers in return for handing over some personal information, for instance information collected through cookies in the web browser or from social media accounts. This might lead to an outcome in which consumers attempt to minimize privacy risks while still trying to obtain as much information as possible. In this case, the amount of information exchanged might be suboptimal.

The higher the number of owners involved, the higher the risk of a socially suboptimal outcome due to data fragmentation. The degree of underinvestment in data collection increases with the strength of the externalities and free-riding by others. A good example is consumer investment in review scores and ratings of services such as hotels and restaurants. Consumers cannot internalize any additional benefits from contributing an additional review or rating. They can easily free-ride on the ratings of others and externalize the costs of these ratings. While natural persons have a right to reject cookies and other tools to collect their personal data in web browsing environments, for example in search engines, very few make use of that right and simply accept cookies because it is the lowest cost solution that enables them to benefit from access to online information sources. The search engine provider aggregates the personal data and can internalize the value of the externalities produced by search requests from individual consumers, for instance aggregating them and using them to generate advertising revenue. These data may have value for many other applications, only some of which are explored by the search engine service provider. In this respect, the absence of a clearly defined transferable ownership right to personal data under the EU GDPR may actually avoid excessive data fragmentation and make it easier for firms to collect personal data.

Resource production costs play an important role in the bargaining outcome and the extent of under-utilization of resources in the general anti-commons model by Schultz et al. (2002). The model assumes increasing marginal costs of production. Resource owners will be willing to contribute to a common benefit up to the point where their marginal costs equal their marginal benefits. One can imagine situations where two data owners decide to split the costs and benefits

on a 50/50 basis in order to overcome their coordination problem and achieve higher benefits than feasible with split data resources. Several examples in Schultz et al. (2002) end up with a 50/50 split Nash outcome. However, that split depends on the underlying production costs functions and market power of the parties. Moreover, in the case of digital data collection increasing marginal costs may not be very realistic. The construction of sensors and algorithms usually entails strong economies of scale driven by fixed investment costs with very low and constant or declining marginal costs. Increasing marginal costs may occur to some extent in data analytics when extracting more information from a dataset may run into increasing costs. Variations in cost structures may lead to Nash bargaining outcomes with corner solutions whereby all benefits accrue to the party with the most advantageous cost structure. For example, in the above case where a consumer trades his personal data for information from a service supplier, the opportunity cost of not receiving the information may be high for the consumer while the opportunity cost of not receiving personal data from one additional person may be very low for the service provider.

While de-fragmentation can overcome the problems of under-use and under-valuation of datasets, Schultz et al. (2002) argue that the transaction costs incurred to achieve integration are usually higher than those associated with the original fragmentation of ownership rights and thus create an asymmetry between the two situations, or between moving from one situation to the other. Their general argument is that private ownership may be good to offset the deadweight losses that come with commons but excessive privatization may lead to greater losses from anti-commons problems. They conclude that fragmentation may be occasionally ex ante efficient given the specific goals of data owners but it may result in inefficient ex post allocations.

Despite these warnings, there are many examples of complex fragmented settings where many parties hold a stake in data collection and use, and were nevertheless able to negotiate contractual arrangements and satisfying solution for all parties involved, despite high transaction costs. Examples range from oil drilling to health and neuroscience data (McPherson et al., 2016; Book et al., 2016). Surveys of firms involved in complex data settings suggest that they are quite satisfied with this negotiated approach (IDC, 2015a). Legislative intervention to allocate some rights to specific parties might reduce the space for negotiations and may result in less satisfactory arrangements for all parties involved. There are many shades of grey in data ownership and the licensing arrangements to access data (Merges, 2008; West, 2007). Fragmented rights across many parties limit the importance of residual ownership rights. In the end it may not be so important anymore who holds that right since the scope of that right is considerably restricted by the contractual arrangement. Realizing the full value of that property right requires collaboration and agreement between many parties and cannot be achieve by a single owner on his own.

Finding a balance between the two extreme solutions of commons and anti-commons depends on the costs and benefits associated with each of these solutions. There is an optimal degree of fragmentation or decentralization of ownership in the Schultz et al. (2002) model. Digitization of data, and the economies of scope in aggregation that it brings, may well have shifted that balance somewhat away from private ownership and towards some degree of aggregation. That would explain the success of digital platforms or multi-sided markets; they facilitate data aggregation and the realization of the benefits from economies of scope in data.

Example 1: Agricultural machinery equipped with digital sensors

Elixson & Griffin (2016) document how agricultural machinery manufacturers try to claim exclusive ownership rights on the data collected by sensors, processors and algorithms inside the equipment, even if that equipment has been bought and is owned by farmers[23]. That results in different ownership structures for the data and the equipment, two complementary tools that jointly produce the output delivered by the machine.

The underlying economics of the farm machinery example can be explained by means of an adapted version of the Ultimatum Game. There are two complementary products in this game, a machine and data. The farmer chooses to buy a digitally equipped machine because it is more productive than a non-digital machine, thanks to the build-in data collection and processing tools. Moreover, the farmer may share the data with his bank in order to obtain better credit conditions. The productivity gains are higher than the benefits from lower interest rates. The manufacturer benefits from economies of scope in data aggregation across all machines from his brand and sells the data to a harvest forecasting firm or for land and environmental management. These uses of aggregated data are beneficial to society. To avoid competition in the harvest data market the manufacturer wants exclusive data ownership that excludes data sharing with the bank. He implements this clause by locking the data in the machine (they are not portable). The farmer is not interested in selling data to a harvest forecaster because transaction costs would be higher than the expected revenue from his single data set.

The farmer has an objective interest in accepting the deal. It increases but does not maximize his revenue from his investment. However, the farmer may reject it because he feels that the deal is not fair: he should have access to the data that he has produced. The manufacturer's sales revenue increases when the farmer accepts the deal. His aggregated data revenue is not affected by one additional unit of data. However, he does not want to create a precedent by allowing the farmer to access to the data because it may unravel his data exclusivity in all other machine sales. Society has an interest in the deal getting accepted. It would trigger a marginal increase in agricultural productivity. The impact of one farmer's data on the accuracy of harvest forecasts is zero. But unravelling of all harvest forecasting would increase uncertainty in agricultural product markets. In short, the farmer's response is crucial to the realization of overall social welfare benefits: will he respond rationally or will he reject the perceived unfairness?

The farmer cannot internalize the benefits from economies of scope from data aggregation across machines. These benefits will be lost for society if the farmer, as owner of the machine, claims exclusive rights to the data. Society may be better off and able to overcome the anti-commons problem in data if the manufacturer retains data ownership rights, unless alternative arrangements can be found that enable the realization of economies of scope in the data. The farmer may sell his data to another data aggregator, provided they are portable, but transaction costs for concluding an alternative deal may be too high compared to the price he receives for his individual data. Separate sensor manufacturing firms may emerge to collect the data at the machine but the cost of retro-

---

[23] In "The end of ownership", Perzanowski & Schultz (2016) explain how the principle of non-exhaustion of copyright at the point of sale of digital products is increasingly applied by producers of copyright-protected products to reject full ownership by the buyer. This legal principle does not only apply to digital books, films and music but also to software and algorithms that collect and process data.

fitting them may be high compared to doing this during the manufacturing process. An intermediate solution would be for the manufacturer and the farmer to accept sharing the data, knowing that it is unlikely anyway that the farmer will sell the data to the harvest forecaster.

This farming equipment example illustrates the economic impact of different data ownership and access options, for the private parties involved as well as for society. Similar mechanism are at work in many other situations where ownership of hardware and data are complementary but fragmented in anti-commons or joined in a wide variety of hybrid semi-commons arrangements that fall short of full common ownership. Shared data ownership arrangements between the manufacturer and the farmer still require negotiation of the specific rights of each party, including the right to re-use and re-sale, and thereby trigger the bargaining situations that do not guarantee a social welfare-maximizing outcome. The impact of these arrangements on overall welfare and on the distribution of welfare between the parties is an empirical question that cannot be solved through theoretical and legal reasoning alone.

Example 2:   Trading personal data with "free" information service providers

The Ultimatum Game can be used to examine the economic impact of bargaining over personal data exchange between a data subject and an information services provider who collects and aggregates data. The data collector makes an opening bid that typically consists of access to information requested by the consumer in return for acceptance of ads and collecting personal data – the standard business model of many "free" online services. If the consumer accepts the bid, the data collector benefits from the sale of targeted ads to advertisers. The consumer gains from access to the information provided. The cost for the consumer consists of the perceived welfare losses from releasing personal data. The latter component is not measurable and often mired in inconsistency due to the "privacy paradox". There is an extensive research literature that estimates consumers' willingness to pay for privacy protection (for an overview, see Acquisti et al., 2015, pp 480-485). Virtually all that research relies on behavioural experiments in artificial settings, not in real life situations. It stumbles into the so-called "privacy paradox" or a gap between consumer's stated preferences and actual behaviour. Acceptance or rejection of the bid hinges on this subjective perception of privacy costs. The ultimatum game can capture this subjective component.

Consumer rejection has impacts for society as well. It may lead to anti-commons problems that may prevent many social welfare benefits from economies of scope in the aggregation of personal data. For example personal data can be valuable for health, environmental and traffic management, for sentiment analysis and security issues. On the positive side, personal data aggregation and analysis drives search engines that help consumers find their product choices. The advertising model enables many other benefits for consumers, such as social networking, access to media products and other free online services, most of which rely on some form of economies of scope in data aggregation. Policy makers walk a thin line between enhancing privacy protection and not losing the social welfare benefits of data aggregation and overcoming anti-commons in data use.

Example 3: Personal Information Management Services

Personal Information Management Services (PIMS) are tools that were designed to enhance the management of personal data rights for data subjects (for instance, see Poikola et al., 2015). PIMS are private service providers that allow individuals to store their personal data. Companies that want to use personal data can approach individuals and ask them for permission to use the data via the PIMS. PIMS do not "sell" or trade personal data but rather facilitate the procedure and provide terms & conditions in line with data protection law. PIMS establish interoperability of personal data between many services. Most importantly, they enable individuals to maintain traceability of personal data and control over any downstream data use. They keep personal data "on a leash" controlled by the data subject. They can retrieve data about their on-line presence, including data from browsing histories, bookmarks, address books, credentials, location and financial data, social network activity, etc. Some PIMS are designed to keep user data in a single place inside the PIMS or to securely transfer them to other services providers. PIMS keep track of the use of personal data, even if they are used by other service providers. They maintain a link between the data subject and the use of their data. They decrease transaction costs for individuals who want to exercise and enforce their specific personal data rights under the EU GDPR, compared to bilateral one-to-one negotiations with each data user (Ctrl+Shift, 2014). In the currently dominant personal data business model on the internet data subjects have no control over further downstream use of their data once they have authorized an internet service provider to collect their personal data, for instance by clicking on a cookie authorization (European Data Protection Supervisor, 2016). Koutroumpis et al. (2016a) argue that decentralized data market places[24] based on tracking systems may be an efficient approach to the management of personal data to address privacy requirements by offering more control to the data subject. On the other hand, PIMS increase transaction costs for data collectors as they have to ask permission for access and use of personal data. This may explain why many data collectors are not enthusiastic adopters of PIMS.[25] As such, PIMS do not overcome the Ultimatum Game bargaining problem between data subjects and data collector firms. The latter may still refuse access to a service if a person is not willing to share his personal data directly rather than through PIMS. That creates a bargaining situation where differences in marginal costs and forgone benefits between both parties will determine the outcome (Schultz et al., 2002). The law may give individuals specific rights but bargaining will determine whether these rights are effectively used or not.

---

[24] Various concepts of personal data markets have been proposed that combine aspects of the centralized and decentralized designs (Bergman et al., 2003 & 2008; Larsen et al., 2015).
[25] However, it is also important to note that the willingness of companies to adopt PIMS may depend on their data collection capabilities. For instance, while large-sized data holding firms currently do not appear to be enthusiastic about PIMS, companies that do not have the same data collection capabilities (and possibly should not be forced to develop them) are more likely to find PIMS beneficial.

5.   Data trade

So far we focused on the legal situation with regard to ownership rights in data and to what extent these legal rights are effective in economic transactions and bargaining situations. In this chapter we move beyond data ownership. We assume that de facto ownership or control rights exist and we explore the possibilities for, and welfare impact of, commercial data trade in these imperfectly defined legal settings.

The starting point for the economics of data trade is the Arrow Information Paradox (Arrow, 1962). In order to sell information the buyer of information must be able to place a value on the information and determine how much she is willing to pay. But once the seller discloses the information, the buyer is in possession of the subject of the trade and no longer has any reason to pay for it. The conventional legal solution to the paradox is a grant of intellectual property rights. If information is subject to a patent or a copyright, then it can be disclosed without fear that it will be taken without compensation. Burnstein (2012) explores several other options outside intellectual property or ownership rights to solve the paradox. He argues that the paradox is not necessarily fully applicable to all types of data. Some data may be more excludable than others, for instance through technical measures. Data can be heterogeneous, split up in packages and displayed in bits. Showing part of the data does not necessarily reveal all the data. That enables a gradual approach to information trade and disclosure. Following Gilson et al. (2009), Burnstein argues that contracts can play an important role in building the institutions that create trust and collaboration in the face of information asymmetries. Contractual relations establish governance mechanisms in lieu of the more familiar risk-allocation provisions of conventional contracts through which the parties engage in mutual information sharing and product development over the course of several years. The parties overcome opportunism by engaging in a collaborative process that build trust and enable the exchange of sensitive and detailed information and raise switching costs of finding another partner, thereby discouraging defection. The accounts from a wide variety of industries, from oil drilling (McPherson et al., 2016) to neuro-imaging (Books, 2016), confirm that this approach "in the shadow of the law" is widely applied. Second, social norms can complement these information exchange contracts. Saxenian's (1996) comparison of technology clusters in Silicon Valley and Route 128 in Massachusetts shows that the critical driver of economic performance in Silicon Valley was an industrial organization that encouraged the free flow of information between firms and conducive to innovation. By contrast, that flow of information and employees between firms was much more restricted in Massachusetts. Burnstein (2012) also mentions data disclosure practices in several innovation driven industries, including in venture capital funding rounds, that are based on a step-wise approach with trust mechanism that lead to gradual disclosure as trust increases and collaboration arrangements become more concrete.

In this chapter we start we "naked" data trading whereby all data are transferred between parties. These are often markets for personal data that are mainly used in advertising. More complex data sets are often not traded in that way. Data owners are conscious of the Arrow Information Paradox and try to design methods that reveal only part of the data and in a gradual profit-maximizing pricing strategy. We then move to multi-sided markets where data are often not traded at all but their information value is leveraged through other mechanisms. We already referred in the previous

chapters to data-driven business models do not require revealing the data in order to commercialize them, for instance in online advertising, a typical activity in a multi-sided market. We conclude this chapter with the observation that much research remains to be done on the role of data in these multi-sided markets.

5.1. Commercial data markets

Some personal data are traded in active markets organised by intermediaries or commercial data brokers. FTC (2014) examines the business models and practices of nine exemplary data brokers. These data brokers compile information on consumers from publicly accessible Internet posts and from their online purchases, browsing history or filling out of warranty cards. They typically collect data from government sources, e.g. the U.S. Census Bureau, other publicly available sources, e.g. social media, blogs and the Internet, and commercial data sources. In addition, FTC (2014) reports three main types of products of data brokers: (1) marketing, e.g. online and offline marketing and marketing analytics, (2) risk mitigation, e.g. via verification of the identity of individuals or fraud detection, and (3) people search, e.g. tracking of activities of competitors or finding old friends.[26] Data brokers typically not only sell raw data, e.g. "actual data elements" such as a person's name, address, age etc., but also "derived data elements" that they infer from the raw data. For instance, data brokers infer a person's interest in a product or service from frequent visits of particular websites or magazine subscriptions. They often also provide "data append" products to help their clients fill in gaps in existing customer information. To illustrate, an important service in online marketing is the so-called onboarding which consists of three subsequent processes: segmentation, matching, and targeting. First, the clients of a data broker ask the broker to identify individuals with particular characteristics (segmentation). Second, the data broker is asked to find these individuals online (matching). Lastly, the matched consumers are targeted online, e.g. via cookies that include additional consumer information appended by the data broker and display ads of the data broker's clients (targeting).[27] This illustrates how economies of scope in data are also at the roots of data brokerage trade. While each data source may provide only a few elements about a consumer's activities, data brokers aggregate these data to create a more comprehensive picture of consumers. FTC (2014) suggests that while data brokers provide their clients with data that may help them to give consumers more choices and lower transaction costs, their business model raises privacy concerns as they compile information about consumers without necessarily directly interacting with them.[28] It is also noteworthy that there are typically multiple layers of data brokers between the raw data and the data element sold to the final clients of a data broker. These multiple layers often make it difficult, if not impossible, for consumers to trace back the data used by the clients of a data broker.

---

[26] According to FTC (2014), there are three aspects to product development by data brokers. First, the creation of data elements to identify consumers with homogenous characteristics and the development of models to predict consumer behaviour. Second, data suppression, e.g. most data brokers refrain from using information about children. Third, data storage, e.g. data brokers store individual data profiles.

[27] See also Bergemann and Bonatti (2011).

[28] See also Bergman et al. (2003, 2007) for thorough analyses of the user-subjective approach of Personal Information Management Systems (PIMS).

In another case study of data markets, Muschalle et al. (2012) stress the new business models and cost saving potential that data sellers, e.g. vendors of data warehouse solutions and algorithm providers, offer their clients (henceforth, data buyers). Similar to FTC (2014), Muschalle et al. (2012) distinguish between raw data ("actual data elements") and analytic results ("derived data elements") as electronically tradable goods. There are two typical scenarios for data markets. First, data buyers use the data market to retrieve indicators which describe the value of a good. These indicators are then correlated with a monetary transaction. Examples for this scenario are marketing measures on certain web forums that effectively influence buying decisions of customers. In a second typical scenario of data markets, data buyers demand factual pieces of information about a given object from different sources to merge them into a single and clean representation of this object. This process, also known as "data fusion", provides customers with a unified view of several multiple data sources (Bleiholder and Naumann, 2008). However, Muschalle et al. (2012) report the following main pricing strategies of data sellers: Freely available data from public authorities, e.g. statistical data[29]; usage based prices; package pricing; flat fee tariffs; two-part tariffs consisting of a fixed fee and a fee per unit sold; and a freemium where the data vendor provides basic services for free while they charge customers for premium services. Finally, according to Balazinska et al. (2011), a crucial challenge for research on data markets is to understand the pricing strategies on data markets (a natural topic for economists)[30] and which effects data and algorithm providers have on the value of data through transformation and integration of data (a natural topic for data scientists).[31] In the following, we focus on the former challenge and outline different models of pricing strategies as proposed in the theoretical economics literature.

5.2. The theoretical literature on data selling

In the previous chapter we discussed how some specific economic characteristics of data (economies of scope, data as intermediary inputs) made them less suitable to fit into the standard economic model of intellectual property rights. Here we discuss some other economic characteristics that affect trade in data.

Unlike ordinary goods, data cannot be inspected by potential buyers. Once the buyer has access to the data he has no incentive to pay for them anymore. Sellers thus need to design partial disclosure procedures. This strand of literature examines the disclosure and sale of information, e.g. revealing product quality (Grossman, 1981; Grossman and Hart, 1980; Milgrom, 1981), selling financial information (Admati and Pfleiderer, 1986, 1990) or disclosure of product variety (Celik, 2014).[32] Data sellers typically have high fixed cost to create their products and services, e.g. costly implementation of IT infrastructure and processing technologies, and low marginal costs. This cost structure, which is

---

[29] While freely available data from third-party sources may not be sold on a data platform as a stand-alone product, it may help to attract customers to paid data services by integrating it with in-house private data.

[30] See the overview of the theoretical economics literature on selling information below.

[31] See also the website of the Data Eco$y$tem project: http://cloud-data-pricing.cs.washington.edu/ (last accessed November 4, 2016).

[32] For models of disclosure where sellers set prices in addition to disclosing horizontal match-value information, see Ottaviani and Prat (2001), Johnson and Myatt (2006), Eső and Szentes (2007), Bergemann and Pesendorfer (2007), and Li and Shi (2013). See also Celik (2014) and Koessler and Skreta (2014).

typical for markets for information goods, drives the pricing strategies in data markets.[33] Typically, an information good is priced according to its value to its consumers, e.g. via versioning, and not as a mark-up on its unit cost. For instance, the data vendors examined in Muschalle et al. (2012, p. 7) "agree that it is almost only the demanders' preferences that determine prices". According to Shapiro and Varian (1999), this value-based pricing typically leads to differential pricing. In a recent paper, Bergemann et al. (2016) study the interaction between a monopolistic provider of information and a single buyer of information where the monopolist has all the relevant information that the buyer needs to solve a decision problem. Their leading example is the market for online advertising where large data holders choose how much information they provide and how they price access to their database, e.g. information about consumer profiles including browsing and searching history. [34] In this case, a contract between a seller and buyer of data specifies which attributes of consumer profiles the seller shall release. In Bergemann et al. (2016), buyers are heterogenous in terms of willingness to pay for additional informative signals as they differ in their prior information on consumer profiles.[35] Bergemann et al. (2016) find that the optimal mechanism consists of at most two information structures suggesting a limited use of versioning.[36] As an idea for further research, Bergemann et al. (2016, p. 23) suggest to examine "the effect of competition among sellers of information (i.e. formalizing the intuition that each seller will be able to extract the surplus related to the innovation element of his database)."

In addition, Bergemann and Bonatti (2012) set-up a three-player model of data provision and data pricing in a Hotelling model with horizontal price differentiation.[37] In their model, a data broker chooses the amount of information and the price to access it for two competing firms. Bergemann and Bonatti (2012) show that the profit-maximizing information policy is to provide only partial information to the competing firms or even adopt exclusive information policies. Bergemann and Bonatti (2012, p. 2) distinguish between three main types of data vendors depending on the source of the data: (1) financial data providers (Bloomberg or Thomson Reuters), including credit rating agencies (Equifax, Transunion, Moody's or Standard & Poor's). (2) Data brokers, e.g. LexisNexis and Acxiom, that compile huge databases on individual consumers. (3) Online aggregators, e.g. Spokeo and Intelius, that mine publicly accessible data to create consumer profiles. Independent of the source of data, these data vendors sell access to their databases and data analytics to downstream firms which in turn use the data to improve their product positioning. As an additional example, Bergemann and Bonatti (2012), consider online platforms such as BlueKai that promote the exchange of user data. In a more recent paper, Bergemann and Bonatti (2015) set up a model of data provision and data pricing with a single provider of a database selling information encoded in cookies on heterogenous consumers to advertisers. These heterogeneous advertisers buy data on individual consumer characteristics in order to tailor their spending on a targeted group of consumers. Bergemann and Bonatti (2015) also study the consequences of competition among data

---

[33] See Shapiro and Varian (1999).

[34] The view that data and information are the same thing is not unanimous. See, for instance, Boisot and Canals (2004).

[35] In the model, the information structures offered by the monopolist are experiments á la Blackwell (1951, 1953). Blackwell's theorem derives conditions under which one information structure, i.e. experiment, is more informative than another. See also Cremer (1982).

[36] See also Shapiro and Varian (1998).

[37] Notably, although Bergemann and Bonatti (2012) discuss a multisided platform as one of their leading examples, they refrain from using multisided platform theory in their theoretical analysis.

sellers that have exclusive information on different consumer segments. They find that the prices of data are higher the higher the fragmentation of data sales which suggests that monopolistic data provision does not necessarily have negative welfare properties. The intuition behind the price-increasing effect of data sales-fragmentation is that exclusive data sellers fail to internalize positive externalities that are present across data sales.

5.3. Data trade in multi-sided markets

Platforms or multi-sided markets represent the latest development in the digital economy. The early stages of the internet and the application of digital technology were characterized by a simple transfer of non-digital offline activities to digital and online environments. Writing texts and sending messages moved from typewriters to computers and online message boards. Shops moved from Main Street to e-commerce. Information costs were reduced, product variety and price competition increased and that brought some changes in market structure. More recently however, platforms have changed the way of doing business online. Platforms can use their data collection and analytics to facilitate matching users on different sides of the market, for example buyers and sellers. "Pure" platforms do not produce goods or content; they only facilitate matching between content suppliers and consumers. Online market places like eBay and Amazon Market Place, and sharing or collaborative economy platforms like Uber and AirBnB, are good examples of platforms. Unlike traditional retailers, they do not buy products for re-sale; they only mediate between sellers and buyers. All this intermediation is driven by data.

Traditional economic models of multi-sided markets (Caillaud & Julien (2003), Rochet & Tirole (2003, 2006), Parker & Van Alstyne (2005), Rysman (2009)) have no explicit role for data collection and use. They revolve around network effects that create a "chicken & egg" problem for commercial strategies to expand the market share of the platform. They are driven by pricing strategies and network effects that drive the number of users on each side of the market. The pricing structure may be distorted to bring on board one side of the market and thereby attract the other side via indirect network effects. Pricing strategies are driven by price elasticities on different sides of the market. Because of network effects and abnormal pricing strategies, that MSM literature became very focused on competition policy issues (see, for instance, Armstrong (2006), Evans (2003) and Rysman (2004)). Data collection, analysis and access play, at best, only an implicit role in these models. In a way, MSM models are now in the situation where neo-classical economic models were three decades ago: information is very important to make the market work but the models do not account for that (imperfect/asymmetric) information.

Thinking on multi-sided markets has evolved towards a focus on data issues. They collect data from all market participants on all sides of the market, aggregate and analyze these data. Because they can aggregate data across users, platforms have a better overview of markets than each of the individual users on the platforms who can only observe their own behaviour. They benefit from economies of scope in data aggregation. The benign view on data collection by platforms is that they use this to facilitate matching users on different sides of a heterogeneous market. Platforms contribute to social welfare because they reduce transaction and information costs that constitute

deadweight losses in society. In that sense platforms are data-driven market innovators[38]. In reality, some of these benefits accrue to platform users; the rest is internalized and monetized by the platform to boost its profits and strengthen its market position. Platforms behave like all profit-maximizing firms and will use the data to benefit their own position. They may drive a wedge between the interests of users on several sides of the market in order to boost their profits, for instance through price and quality discrimination in their matching algorithms. They may tweak the matching mechanism in their own favour, subject to limits imposed by user behaviour, and thereby introduce a new source of information asymmetry and transaction costs. A rapidly growing literature on search rankings investigates these questions (Ursu, 2015; Farronato, 2015) though there is no explicit role for data in these rankings. Platforms may extract data rents from market users in various ways, making some data accessible in return for changes in the pricing structure, or hiding the data to safeguard their asymmetric rents. This raises questions about data ownership & access, and the impact of such strategies on social welfare.

The trade-off between upstream and downstream data-driven innovation, as discussed in section 3.1) takes a new turn in multisided platforms. A platform owner owns the platform data and may give downstream innovators access to (part of) the data as an incentive to join the platform. For example, firms operating on the Facebook platform can make use of the social graph data that Facebook collects for marketing purposes. Similarly, Google Search has put in place an ad auction mechanism that enables advertisers to benefit from the search terms data that it collects – though without direct access to the search terms.

Since there are competing platforms, the platform owner has an incentive to maximize innovation on his platforms and offer generous conditions to complementary data producers (see Gawer & Cusumano for the dynamics of technology platforms). For example, mobile app platforms are open to innovators who can use it to distribute their apps, in return for access to the data that they collect. User access to the platform as well as the platform's access to the data generated by the users often comes with technology standards that define the flow of data inside platforms. Usually there is no margin for negotiation, especially in the case of small users on large platforms. West (2005/2007) discusses the drivers of specification rights (for technology standards) that equally apply to access conditions (for data sharing).

The gap between existing economic models of multi-sided markets that have no explicit role for data ownership, use and access, and the reality of these markets that is entirely driven by data, persist.

---

[38] See also OECD (2015) on data-driven innovation.

41

6. Data access and portability

6.1. Open versus closed data

Private data ownership rights should not be confused with open or closed access. Privately owned property can be open access while non-proprietary assets can be de facto closed for access (Merges, 2008). The question of whether or not data should be openly accessible (purely open access to data) is a debated issue in academia (Dewald et al., 1986; Glandon, 2011; McCullough et al., 2006) and policy (European Commission, 2012; ESRC, 2010; OECD, 2007). Advocates of open access to data argue that it facilitates subsequent research, including replication of existing works, and increases the diffusion of knowledge thereby enhancing the efficiency of the research system (Begley and Ellis, 2012; McCullough et al., 2008; Mueller-Langer et al., 2017; Nature, 2009; Piwowar et al., 2007; Piwowar and Vision, 2013).

Recent research estimates the direct financial benefits to society of open access to public or government data. According to the Danish Enterprise and Construction Authority (2010), the direct financial benefits of open access to Danish address data totaled EUR 62 million while the costs of the underlying agreement have been approximately EUR 2 million.[39] Loomis et al. (2015) conducted a survey among the 44,731 data users registered with the U.S. Geological Survey to examine their willingness -to-pay for publicly available Landsat data, e.g. satellite images of the earth. Using the contingent valuation method to valuate Landsat data, Loomis et al. (2015) calculate the annual economic benefit of Landsat via its open data policy for the U.S. to be $1.8 billion in 2011. However, individual incentives of data creators and societal interests are not always well aligned. Recent empirical evidence suggests that data creating researchers typically do not voluntarily share their data (Andreoli-Versbach and Mueller-Langer, 2014). There are many reasons why data creators are reluctant to share their data. First, data creation is often costly (Fecher et al., 2015).[40] Second, data creators shall protect the competitive advantage that is associated with the data (Mueller-Langer and Andreoli-Versbach, 2014). Third, data creators have an incentive to fully exploit the private value of the data (Stephan, 1996). However, purely private ownership of data ("anti-commons") may diminish access for downstream innovators and users of the data and may create losses for society (Loomis et al., 2015). An intermediate solution would be "semi-commons" whereby some parties have some use and access rights to the data and others hold other rights (Fennell, 2011; Heverly, 2003; Bertacchini et al., 2009). In complex settings where many parties hold a stake in data collection, transformation and use complex contractual arrangements may emerge that fragment the rights and obligations across many parties. There are many examples where markets have worked out such complex arrangements, ranging from oil drilling to health and neuroscience data. Surveys of firms involved in complex data settings suggest that they are quite satisfied with this negotiated approach (IDC and Open Evidence, 2015a&b). Legislative intervention to allocate some rights to specific parties might reduce the space for negotiations and may result in less satisfactory

---

[39] See also Lind (2010).

[40] However, there also instances where data is created as a by-product of transactions with low or zero marginal cost.

arrangements for all parties involved. Enabling the exploration of aggregated data for the benefit of society implies that contractual arrangements to enable access to the data should be feasible, independently of the ownership structure and not hindered by legal constraints. There are many shades of grey in data ownership and the licensing arrangements to access data (Merges, 2008; West, 2007). The law can allocate an ownership right to one party but the eventual use and ownership of that right will depend on market forces. The owner may well decline to use it, in his own interest.

6.2. Data portability and interoperability

The notion of data portability has attracted a lot of attention lately. It refers to data subjects' option to move their data between different (hardware and/or software) environments. In this case, a data subject can be an individual or a firm. In the first case, portability refers to personal data, while in the second case it refers to non-personal data. However, the legal and the economic implications of portability differ. In the recently adopted GDPR, the concept is defined as a right of data subjects to have their data extracted from a data controller (or service provider) to be able to move it to a different data controller and/or simply store it. From an economics perspective, however, data subjects will be willing to port their data only if doing so represents a net gain (less cost or more satisfaction, for instance). The GDPR contemplates the option of having the data transmitted directly from one controller to another, when technically feasible. Hence, while there is no obligation for data controllers to make this happen, in order to reach real efficiencies, data portability requires system interoperability. In this context, interoperability refers to the ability of two or more systems (online service providers or platforms) to exchange information and/or data, and subsequently be able to use it.

In a standard market setting, portability lowers switching costs (Klemperer, 1995), i.e. it reduces the cost to move from one data controller to a functionally identical online service supplied by another firm. That, in turn, should promote competition between these online services. This benefits all subjects, not only those that decide to take their data to competing service providers, by means of increased choice, lower prices and/or better quality . Competition benefits users but not necessarily the service providers who would prefer to have many users locked into their service by lowering entry costs and increasing exit costs. Data portability means more competition -as services will become closer substitutes- as well as higher costs, since portability requires not only implementation costs but also usage costs. In addition, data portability could also imply adopting different strategies or changing business models, for instance, offering better quality or customer care to avoid excessive churn. This standard single-sided market conclusion may change in the case of online platforms that have the characteristics of multi-sided markets (Rochet and Tirole, 2003): they bring together suppliers and consumers, and possibly other categories of users (like advertisers, app developers and other service providers, etc.). The main feature of these platforms is the presence of indirect network effects between these user groups: what happens in one user group affects the other user groups as well.

Although these data portability concerns related to platforms share similarities with number portability as introduced in the EU telecommunications sector some years ago (Graef, 2015a), there are no examples related to digital technologies. Email address portability, for instance, has been highlighted as a possible starting point . Despite the similarities, there are also relevant differences. Although some users multi-home, telecom services are basically substitutes: for the majority of users, a single service suffices. Hence, when a user wants to change a supplier, there are efficiencies in preserving the same number, enabled by number portability. However, most information society services offer differentiated products, and hence are of a more complementary nature. In most real-world platform markets, there are several competing firms and at least one-side appears to multi-home (Caillaud and Jullien, 2003). For instance, consumers can indistinctly use ebay or Amazon to purchase products and at the same time, and many sellers are active in different marketplaces. In this case, both sides multi-home. Here, data is context-specific, and may not be valuable to port it to a different platform. As an example, consider an individual that uses Facebook to connect with friends and family and LinkedIn to connect with professional colleagues. She may find little value in moving her data from one social network to the other, since she is using the different platforms for different purposes. In the same line of argument, business users of online advertising services may use different providers to focus on different market segments. For instance, app developers can in principle multi-home, i.e., they can develop versions of their apps that can be installed in different operating systems. Hyrynsalmi et al. (2016) assessed multi-homing in mobile application ecosystems. Although multi-homing can be a substitute for data portability (Doganoglu and Wright, 2006), each market will differ and no standard rule fits them all.

The link between switching costs (data portability) and network effects is that they both lead users to value interoperability, though with slightly different meanings and implications. In the presence of switching costs, interoperability is understood as the ability for a user to take advantage of the same investment between her own purchases. The value users place on such ability is measured by the premium they are ready to pay for keeping the same product/service, or for staying with the same vendor. With network effects, interoperability is understood as the ability to communicate directly with, or take advantage of the same complements as other consumers. Here, the value placed on interoperability translates in an increase in the user's willingness-to-pay to join larger networks. As a result, users often face lock-in, resulting either from their own previous choices or from other users' choices. Such lock-in confers a potential ex-post power to platforms: in the presence of switching cost, the price structure defined by the platform would penalise the user side with higher switching costs, and will be able to fix prices above costs by an amount equal to the user's switching cost.; in the presence of network effects, the larger the network the larger the price the platform can charge to the users.

In an online platforms environment, there are other forces that will operate along with switching costs in the determination of the platform market's competitive intensity (Duch-Brown, 2016). These other factors are: the presence of economies of scale; users' preferences that allow for platform differentiation; the platform's capacity (or congestion) constraints; direct network effects and; especially, indirect network effects. An example is online marketplaces. In this case, indirect network effects are strong: buyers benefit from a large number of potential sellers and sellers benefit from large number of potential buyers. Although switching costs are low for buyers who can easily multi-home, i.e., buy from competing marketplaces, these are particularly high for sellers. For them, their selling rating (reputation) can be considered as a specific investment since it depends on the number

of transactions a seller has successfully completed on a given platform. If sellers would like to move to a different platform due to better services or more favourable fees, they would like to take their data with them. Additionally, it is difficult to build up reputation on several platforms, as reputation depends on the number of transactions a seller has already honestly completed on a given network. Transferring reputation from one platform to another is rather difficult or often even impossible. Hence, investment into one's reputation is typically platform specific. However, even if data portability is allowed, indirect network effects could still imply that the move to a different marketplace is not profitable: it will depend on the volume of potential buyers that are using the alternative marketplace. In two multi-sided markets, sellers could be locked-in not only because restrictions on data portability that increase switching costs, but also because the strength of indirect network effects (Casadesus-Masanell and Ruiz-Aliseda, 2009).

Under full platform interoperability, users from one side would be able to interact with all the users on the other side, regardless of the platform they have originally adopted. As a result, users – irrespective of their side- would not benefit from platform specific network effects and would perceive all platforms as identical in that they would give access to the same pool of users in all sides. For instance, if a Facebook user can post comments or upload photos in a friend's Google+ account, the incentives to switch platforms are heavily reduced, and she will perceive the platforms as being identical. Platform homogeneity could lead to a fierce competition to attract fee-paying users to increase revenue. However, users would have only weak incentives to switch since they would reach all users on different sides from their original platform. As an example, consider an advertiser that can reach users of all social networks from its original platform. Competition for advertisers will tend to equalise prices and, at the same time, will increase the number of potential viewers. In this case too, incentives to switch platform may be significantly reduced. Mobile apps can be used as a final example. Different from the Internet, the mobile app world is quite fragmented. First, users need to download the apps they want to use. Once in the app, it is simply not possible to navigate to another app, and definitely not possible to use it at all if it has not been downloaded previously. To overcome this limitation, some players in the industry are developing the concept of app streaming, i.e., developing interoperability between apps. The concept is relatively new, but has the potential to completely transform the way both users and developers engage with mobile content and services (Duch-Brown, 2016).

The absence of interoperability can, on the other hand, fragment the market and result in less competition. Consumers would be locked-in and would face costs for switching to alternative service providers, in case these would be substitutes. Alternatively, they would multi-home in the case of complementary services, and interoperability would be achieved through indirect mechanisms (Doganoglu and Wright, 2006). However, given the multi-dimensional nature of the effects at play, many different results can be obtained, and more research is needed to disentangle the forces at play and its effects. For instance, when multi-homing, some individuals will use simultaneously two platforms, increasing each platform revenues and data collection. This provides an incentive to platforms to block interoperability, allowing them to keep some market power. Hence, it is the quest for market dominance that prevents systems from agreeing on interoperability. This happens in traditional and in platform markets. However, in this last case tipping may be driven by user subsidisation on one side of the market (Casadesus-Masanell and Ruiz-Aliseda, 2009).

Characteristics such as design, management, and operation rules, among others, usually differ between service providers. As a result, data subjects will face some switching costs if they decide to use another platform as they have to get used to the terms on the new platform, even if portability and interoperability are guaranteed. In addition, many service providers try – sometimes hard – to create endogenous switching costs in order to bind customers. For instance, data collection by some online marketplaces regarding payments/addresses/buyers' characteristics enable one-click shopping which may increase switching costs for those who value that option. In addition, the increasing importance of mobile commerce can also imply that, with limited device capacity, users may not want to install many shopping apps, only the preferred ones, which can also affect switching costs (Duch-Brown, 2016).

In principle, a welfare-maximizing policy maker would prefer interoperable services in both traditional and platform markets. However, social welfare comparisons are ambiguous given the trade-offs between the benefits of network effects, the costs of entry, and the intensity of users' preferences for platforms (Casadesus-Masanell and Ruiz-Aliseda, 2009). Many of these results are extracted from analyses that do not take data considerations explicitly. Hence, more research is needed to study these issues in an appropriate framework.

.

7.   Conclusions

Economists are inclined to think that well-defined private property rights are a necessary condition for an efficient resource allocation. The question in this paper is to what extent this holds for non-rival data. Would a better specification of data property rights improve efficiency and reduce data market failures? If so, which party should receive these rights and does it matter who receives them?

The case for legal ownership rights to data starts from the Arrow Information Paradox: once data are shown to a potential customer they can no longer be sold because the customer already has the information. Data are non-rival and non-excludable and there is no incentive to invest in their production and commercialization without legal protection of the intellectual property rights type that makes them excludable. Despite gaps in the legal regime, bargaining in data markets produces a de facto ownership or residual rights allocation, both in commercial B2B and in personal B2C data settings. However, bargaining may not produce an optimal social welfare maximizing ownership allocation. Market failures caused by externalities, strategic behavior and transaction costs are likely to occur.

Chapter 2 started with the basic technical characteristics of data and information and derived some economic characteristics from these basic technical features. In chapter 3 we focused on the legal situation with regard to ownership rights in data. Despite the rapidly growing volume and economic importance of data in the digital economy, the legal framework for data ownership, access and trade remains incomplete and often ambiguous, both for firms' commercial data and for individuals' personal data, in the EU and elsewhere. The absence of personal data ownership rights in law does not prevent the emergence of de facto ownership. De facto data ownership seems to dominate. We link the legal debate on data ownership with relevant branches of the economics literature,

including intellectual property rights economics, the commons and anti-commons literature, models of trade under asymmetric information and multi-sided markets.

Chapter 4 explored to what extent these limited legal rights affect economic welfare for the private parties concerned and for society as a whole. Fragmented ownership or anti-commons leads to bargaining between data holders and their potential clients "in the shadow of the law". Anti-commons inhibits the realization of the full economic benefits of economies of scope in non-rival data. It may slow down innovation and affect the efficiency of data markets. For consumers, the ability to reject personal data sharing proposals may be welfare-enhancing in terms of perceived benefits of privacy protection. However, less data sharing, both in B2C and in B2B situations, may reduce social welfare when it increases data fragmentation in anti-commons and blocks economies of scope in data aggregation. The outcomes of bargaining over data ownership and access rights matter for the economic impact of data. The outcomes are not necessarily welfare optimizing.

Chapter 5 moved beyond data ownership. We assumed that de facto ownership or control rights exist and explored the possibilities for, and welfare impact of, commercial data trade in these imperfectly defined legal settings, starting from Arrow's Information Paradox. Intellectual property-like legal protection is only one solution to overcome that Paradox. Contracts and social norms can complement or even substitute for legal ownership rights, and many real-life examples show that they often do so effectively. Besides, the data economy has developed business models that enable data owners to commercialize the value of their data without revealing the data, especially in multi-sided markets or platforms. However, the gap between existing economic models of multi-sided markets that have no explicit role for data ownership, use and access, and the reality of these markets that is entirely driven by data, persist.

What can public regulators do about this? Would a better specification of the scope of data ownership rights improve efficiency and reduce data market failures? Because it is difficult to know ex-ante what the social welfare maximizing arrangement would be, regulators may have little guidance for an intervention. They may decide in favour of a hands-off approach and leave it to the private parties and market competition to work out an arrangement, or probably a variety of contractual arrangements. That is not efficient either because, as explained above, the outcome of private bargaining is likely to be Pareto-inferior and generate welfare losses. Moreover, the bargaining outcome may deliver a very unequal distribution of benefits that depends on production cost structures, including transaction costs.

Allocating data ownership rights to one or the other party may not be welfare-maximizing either. The solution proposed by some German law scholars (Wiebe, 2016b; Zech, 2016) to allocate ownership rights to the party that contributes most to the value to the dataset is hard to implement in a world with strong externalities where individual efforts are subject to free-riding by others. Well-defined ownership rights may increase data fragmentation and prevent the realization of economies of scope due to high transaction costs. Economists (OECD, 2015) and lawyers (for instance, Drexl, 2016) have argued that opening access to data, or piercing ownership rights with à la carte or more general access provisions can solve these problems. Our tentative conclusion is that there are no easy answers to these questions. We offer no policy solutions yet and more research is required to bring economics up to speed with these questions.

Bibliography

Acquisti, A., C. Taylor and L. Wagman (2016), The economics of privacy, Journal of Economic Literature, 54(2), 442–492.

Admati, A., and P. Pfleiderer (1986), A monopolistic market for information, Journal of Economic Theory, 39(2), 400–438.

Admati, A., and P. Pfleiderer (1990), Direct and indirect sale of information, Econometrica, 58(4), 901–928.

Andreoli-Versbach, P. and F. Mueller-Langer (2014), Open access to data: An ideal professed but not practiced, Research Policy, 43(9), 1621–1633.

Anton, J. and G. Biglaiser (2010), Compatibility, interoperability and market power in upgrade markets, Economics of Innovation and New Technology, 19(4), 373–385.

Armstrong, M. (2006), Competition in two-sided markets, The RAND Journal of Economics, 37(3), 668–691.

Arrow, K. (1962), Economic welfare and the allocation of resources for invention, in: The rate and direction of inventive activity (editor: NBER), 609.

Arthur, W. B. (2009), The nature of technology: What it is and how it evolves, Free Press.

Balazinska, M., B. Howe and D. Suciu (2011), Data markets in the cloud: An opportunity for the database community, PVLDB, 4(12), 1482–1485.

Bechtold, S. (2004), Digital rights management in the United States and Europe, American Journal of Comparative Law, 52, 323–382.

Begley, C. G. and L. M. Ellis (2012), Drug development: Raise standards for preclinical cancer research, Nature, 483, 531–533.

Bergemann, D. and A. Bonatti (2011), Targeting in advertising markets: Implications for offline vs. online media. Rand Journal of Economics, 42(3), 417–443.

Bergemann, D. and A. Bonatti (2012), Markets for data, https://economicdynamics.org/meetpapers/2012/paper_538.pdf

Bergemann, D. and A. Bonatti (2015), Selling cookies, American Economic Journal: Microeconomics, 7(3), 1–37.

Bergemann, D., A. Bonatti and A. Smolin (2014), Selling experiments: Menu pricing of information, Cowles Foundation Discussion Paper No. 1952.

Bergemann, D. and M. Pesendorfer (2007), Information structures in optimal auctions, Journal of Economic Theory, 137(1), 580–609.

Bergman, O., R. Beyth-Marom and R. Nachmias (2003), The user-subjective approach to personal information management systems, Journal of the American Society for Information Science and Technology, 54(9), 872–878.

Bergman, O., R. Beyth-Marom and R. Nachmias (2008), The user-subjective approach to personal information management systems design: Evidence and implementations, Journal of the American Society for Information Science and Technology, 59(2), 235–246.

Bertacchini, E., J. P. B. De Mot and B. Depoorter (2009), Never two without three: Commons, anticommons and semicommons, Review of Law & Economics, 5(1), 163–176.

Besen, S. and L. Raskind (1991), An Introduction to the Law and Economics of Intellectual Property. The Journal of Economic Perspectives, 5(1), 3–27.

Blackwell, D. (1951), The comparison of experiments, in: J. Neyman, ed., Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, 93–102.

Blackwell, D. (1953), Equivalent comparison of experiments, The Annals of Mathematical Statistics, 24, 265–272.

Blake, T., C. Nosko and S. Tadelis (2015), Consumer heterogeneity and paid search effectiveness: A large-scale field experiment, Econometrica, 83 (1), 155–174.

Bleiholder, J. and F. Naumann (2008), Data fusion, ACM Computing Surveys, 41(1), 1–41.

Boisot, M. and A. Canals (2004), Data, information and knowledge: Have we got it right?, Journal of Evolutionary Economics, 14(1), 43–67.

Book, G. A., M. C. Stevens, M. Assaf, D. C. Glahn and G. D. Pearlson (2016), Neuroimaging data sharing on the neuroinformatics database platform, Neuroimage, 124(Part B), 1089–1092.

Buchanan, J. and Y. Yoon (2000), Symmetric tragedies: commons and anti-commons. Journal of Law and Economics, 43(1), 1–13.

Bühler, S., R. Dewenter and J. Haucap (2006), Mobile number portability in Europe, Telecommunications Policy 30(7), 385-399.

Cabral, L. (2009), Small switching costs lead to lower prices, Journal of Marketing Research 46(4), 449-451.

Caillaud, B. and B. Jullien (2003), Chicken and egg: Competition among intermediation service providers, RAND Journal of Economics 34(2), 521-552.

Casadesus-Masanell, R. and F. Ruiz-Aliseda (2009), Platform competition, compatibility, and social efficiency. HBS working paper 09-058.

Celik, L. (2014), Information unraveling revisited: disclosure of horizontal attributes, Journal of Industrial Economics, 62(1), 113–136.

Chorost, M. (2011), The World Wide Mind: The coming integration of humanity, machines, and the internet, Simon and Schuster, NYC.

Coase, R. H. (1960), The problem of social cost, Journal of Law and Economics, 3, 1–44.

Cremer, J. (1982), A simple proof of Blackwell's comparison of experiments theorem, Journal of Economic Theory, 27, 443–493.

Ctrl+Shift (2014), Personal Information Management Services: An analysis of an emerging market, Ctrl+Shift, UK.

Cusumano, M. and A. Gawer (2002), The elements of platform leadership, MIT Sloan Management Review: MIT's journal of management research and ideas, 43 (3), 51–58.

Danish Enterprise and Construction Authority (2010), The value of Danish address data, Danish Enterprise and Construction Authority, Copenhagen.

Dewald, W. G., J. G. Thursby and R. G. Anderson (1986), Replication in empirical economics: The Journal of Money, Credit and Banking Project, American Economic Review, 76(4), 587–603.

Doganoglu, T. and J. Wright, (2006), Multihoming and compatibility, International Journal of Industrial Organization, 34, 45–67.

Drexl, J. (2016), Designing competitive markets for industrial data: Between propertisation and access, Max Planck Institute for Innovation and Competition Research Paper No. 16-13.

Drexl, J., R. M. Hilty, .L. Desaunettes, F. Greiner, D. Kim, H. Richter, G. Surblytė and K. Wiedemann (2016a), Ausschließlichkeits- und Zugangsrechte an Daten, GRUR Int., 914.

Drexl, J., R. M. Hilty, L. Desaunettes, F. Greiner, D. Kim, H. Richter, G. Surblytė and K. Wiedemann (2016b), Position Statement of the Max Planck Institute for Innovation and Competition of

August 16, 2016 – On the current debate on exclusive rights and access rights to data at the European level.

Dubé, J. P., G. J. Hitsch, P. E. Rossi (2009) Do switching costs make markets less competitive? Journal of Marketing Research, 46(4), 435-445.

Duch-Brown, N. (2016), The competitive landscape of online platforms. JRC, mimeo.

Ellixson, A. and T. Griffin (2016), Farm data: Ownership and protections. University of Maryland and Kansas State University Agricultural Economics Department, mimeo.

Economic and Social Research Council (2010), ESRC research data policy. Retrieved from http://www.esrc.ac.uk/about-esrc/information/data-policy.aspx.

Economides, N. (1989), Desirability of compatibility in the absence of network externalities, The American Economic Review, 79(5), 1165-1181

European Commission (2012), Towards better access to scientific information: Boosting the benefits of public investments in research, European Commission: Brussels.

European Commission (2017a), Commission Staff Working Document on the free flow of data and emerging issues of the European data economy accompanying the Communication "Building a European data economy", SWD(2017) 2 final, European Commission: Brussels.

European Commission (2017b), Communication "Building a European data economy", COM(2017) 9 final, European Commission: Brussels.

European Court of Justice (2005), Judgments of the Court in Cases C-46/02, C-203/02, C-338/02 and C-444/02 on the protection of databases.

European Data Protection Supervisor (2016), Opinion 9/2016, EDPS opinion on personal information management systems, Brussels.

European Data Protection Supervisor (2014), Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy, Preliminary opinion of the EDPS, March 2014.

European Union (1996), Directive 96/19EC of the European Parliament and the Council, 11 March 1996, on the legal protection of databases. Official Journal of the European Communities, L77/20.

European Union (2016), Regulation 2016/679 of the European Parliament and the Council, 27 April 2016, on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), Official Journal of the European Communities, L119/1.

European Union (2016), Directive 2016/943 of the European Parliament and the Council, 8 June 2016, on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure. Official Journal of the European Communities L157/1.

Eső, P. and B. Szentes (2007), Optimal information disclosure in auctions and the handicap auction, Review of Economic Studies, 74(3), 705–731.

Evans, D. (2003), The antitrust economics of multi-sided platform markets, Yale Journal on Regulation, 20, 325–381.

Evans, D. and R. Schmalensee (2007), The industrial organization of markets with two-sided platforms, Competition Policy International, 3(1), 151–179.

Federal Trade Commission (2014), Data brokers: A call for transparency and accountability, FTC, Washington, DC.

Fecher, B., S. Friesike and M. Hebing (2015), What drives academic data sharing?, PLoS ONE, 10(2), e0118053.

Fennell, L. A. (2011), Commons, anti-commons, semi-commons, Research Handbook on the Economics of Property Law. Ayotte, K. and H. E. Smith (eds.), 35–56, Edward Elgar: Cheltenham.

Gell-Mann, M. (1996), The quark and the jaguar: Adventures in the simple and the complex, St. Martin's Griffin, NYC.

Gilson, R. J, C. F. Sabel and R. E. Scott (2009), Contracting for innovation: Vertical disintegration and interfirm collaboration, Columbia Law Review, 109(3), 431–502.

Glandon, P. (2011), Appendix to the report of the editor: Compliance project report on the American Economic Review data availability compliance project, American Economic Review Papers and Proceedings, 101(3), 659–699.

Graef, I. (2015a), Mandating portability and interoperability in online social networks: Regulatory and competition law issues in the European Union, Telecommunications Policy, 39(6), 502–514.

Graef, I. (2015b), Market definition and market power in data: The case of online platforms. World Competition Law and Economics Review, 38(4), 473–506.

Grossmann, S. and O. Hart (1986), The costs and benefits of ownership: A theory of vertical and lateral integration, Journal of Political Economy, 94(4), 691–719.

Hagiu, A. and J. Wright (2015a), Marketplace or reseller?, Management Science, 61(1), 184–203.

Hagiu, A. and J. Wright (2015b), Multi-sided platforms, International Journal of Industrial Organization, 43, 162–174.

Heller, M. A. (1998), The tragedy of the anticommons: Property in the transition from Marx to markets, Harvard Law Review, 111, 621–688.

Heller, M. A. and R. S. Eisenberg (1998), Can patents deter innovation? The anticommons in biomedical research, Science, 280, 698–701.

Heverly, R. A. (2003), The information semicommons, Berkeley Technology Law Journal, 18(4), 1128–1189.

Hörner, J. and A. Skrzypacz (2012), Selling Information, Discussion Paper CFDP 1743RR, Cowles Foundation, Yale University.

Holldobler, B. and E.O. Wilson (1990), The ants, Harvard University Press, Cambridge, MA.

Hugenholtz, B. (2016), Europe's sui generis database right, in: S. Frankel and D. Gervais (eds.), The Internet and the emerging importance of new forms of intellectual property, Kluwer Law.

Hyrynsalmi, S., A. Suominen, and M. Mäntymäki (2016), The influence of developer multi-homing on competition between software ecosystems, The Journal of Systems and Software, 111, 119–127.

IDC and Open Evidence (2015a), European data market, SMART 2013/0063, D6, First Interim Report, European Commission, DG CONNECT.

IDC and Open Evidence (2015b), European data market, SMART 2013/0063, D 3.6, Data Ownership and Access to Data – Key Emerging Issues, Third Release. European Commission, DG CONNECT.

Ji, S., P. Mittal and R. Beyah (2016), Graph data anonymization, de-anonymization attacks and de-anonymizability auantification: A survey, IEEE Communications and Surveys, 99, 1.

Johnson, J. and D. Myatt (2006), On the simple economics of advertising, marketing, and product design, American Economic Review, 96(3), 756–784.

Katz, M. L. and C. Shapiro (1985), Network externalities, competition, and compatibility, American Economic Review, 75(3), 424-440.

Kerber, W. (2016), A new (intellectual) property right for non-personal data? An economic analysis, Gewerblicher Rechtsschutz und Urheberrecht, Internationaler Teil, 11, 989–998.

Klemperer, P. (1995), Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade, Review of Economic Studies, 62 (4), 515-539.

Koessler, F. and V. Skreta (2014), Mechanism selection by an informed seller: Product and taste heterogeneity, Discussion paper, PSE and UCL.

Koutroumpis, P., A. Leiponen and L. Thomas (2016a), The unfulfilled potential of data marketplaces, mimeo.

Koutroumpis, P., A. Leiponen and L. Thomas (2016b), The economics of data goods, mimeo.

Lambrecht, A. and C. Tucker (2015), Can big data protect a firm from competition?, mimeo.

Landes, W. and R. Posner (2003), The economic structure of intellectual property law, Harvard University Press, Cambridge, MA.

Landsberger, H. A. (1958), Hawthorne revisited, Cornell University, Ithaca, NY.

Larsen, R., G. Brochot, D. J. Lewis, F, Eisma and J. Brunini (2015), Personal data stores, Cambridge Judge Business School.

Laudon, K. C. (1996), Markets and privacy, Communications of the ACM, 39(9), 92–104.

Laudon, K. C. (1997), Extensions to the theory of markets and privacy: Mechanics of pricing information, New York University Stern School of Business Working Paper IS-97-4.

Lemley, M. and C. Shapiro (2007), Patent holdup and royalty stacking. Texas Law Review, 85, 1–59.

Levitt, S. D. and J. A. List (2011), Was there really a Hawthorne effect at the Hawthorne plant? An analysis of the original illumination experiments, American Economic Journal: Applied Economics, 3(1), 224–238.

Lewis, R. and J. Rao (2015), The unfavorable economics of measuring the returns to advertising, Quarterly Journal of Economics, 130(4), 1941–1973.

Li, H. and X. Shi (2013), Discriminatory information disclosure, Discussion paper, University of British Columbia and University of Toronto.

Lind, M. (2010), The value of Danish address data: Social benefits from the 2002 agreement on procuring address data etc. free of charge. Danish Enterprise and Construction Authority.

Loomis, J., S. Koontz, H. Miller and L. Richardson (2015), Valuing geospatial information: Using the contingent valuation method to estimate the economic benefits of Landsat satellite imagery. Photogrammetric Engineering & Remote Sensing, 81(8), 647–656.

Martens, B. (2016), An economic policy perspective on online platforms, JRC Technical Report 2016/05.

Matutes, C. and P. Regibeau (1988), Mix and match: Product compatibility without network externalities, RAND Journal of Economics, 9(2), 221-234.

Mayer-Schonberger, V. (2011), Delete: the virtue of forgetting in the digital age, Princeton University Press.

McCullough, B. D., K. A. McGeary and T. D. Harrison (2006), Lessons from the JMCB Archive, Journal of Money, Credit and Banking, 38(4), 1093–1107.

McCullough, B. D., K. A. McGeary and T. D. Harrison (2008), Do economics journal archives promote replicable research? Canadian Journal of Economics, 41(4), 1406–1420.

McPherson, J., J. de Wardt, M. L. Laing and N. Zenero (2016). Data ownership for drilling automation: Managing the impact. Society of Pretroleum Engineers.

Melnik, M. I. and J. Alm (2002), Does a seller's ecommerce reputation matter? Evidence from ebay auctions. The Journal of Industrial Economics, 50, 337–349.

Merges, R. P. (2008), IP rights and technological platforms. UC Berkeley, mimeo.

Monahan, T. and J. A. Fisher (2010), Benefits of 'observer effects': Lessons from the field, U.S. National Library of Medicine, National Institutes of Health.

Mueller-Langer. F. and P. Andreoli-Versbach (2014), Open access to research data: Strategic delay and the ambiguous welfare effects of mandatory data disclosure, Max Planck Institute for Innovation & Competition Research Paper No. 14−09.

Mueller-Langer, F., B. Fecher, D. Harhoff and G. G. Wagner (2017), The economics of replication, Max Planck Institute for Innovation and Competition Research Paper No. 17−03.

Muschalle, A., F. Stahl, A. Loser and G. Vossen (2012), Pricing approaches for data markets, mimeo.

Nash, J. (1950), Equilibrium points in n-person games, Proceedings of the National Academy of Sciences, 36(1), 48–49.

Nash, J. (1951), Non-cooperative games, The Annals of Mathematics, 54(2), 286–295.

OECD (2015), Data-driven innovation: Big data for growth and well-being, OECD, Paris.

OECD (2016), Maximizing the economic value of data: Understanding the benefits and challenges of enhanced data access, OECD, Paris.

Ottaviani, M. and A. Prat (2001), The value of public information in monopoly, Econometrica, 69(6), 1673–1683.

Palfrey J. and U. Gasser (2012), Interop: The promise and perils of highly interconnected systems, Basic Books, NYC.

Parker, G. and M. Van Alstyne (2014), Innovation, openness and platform control, mimeo.

Pilaszy, I. and D. Tikk (2009), Recommending movies: even a few data are more valuable than metadata, Proceedings of the 2009 ACM Conference on Recommender Systems, 93–100.

Piwowar, H. A., Day, S. Roger and D. B. Fridsma (2007), Sharing detailed research data is associated with increased citation rate, PLoS ONE, 2(3), e308.

Piwowar, H. A. and T. J. Vision (2013), Data reuse and the open data citation advantage, PeerJ, 1, e175.

Perzanowski, A. and A. Schultz (2016), The end of ownership: Personal property in the digital economy, MIT Press.

Poikola, A., K. Kuikkaniemi and H. Honko (2015), MyData: A Nordic model for human-centered personal data management and processing, Finnish Ministry of Transport and Communications, Helsinki.

Rochet, J. C. and J. Tirole (2003), Platform competition in two-sided markets, Journal of the European Economic Association, 1(4), 990–1029.

Rochet, J. C. and J. Tirole (2006), Two-sided markets: a progress report, Rand Journal of Economics, 37(3), 645–667.

Rosen, S. (1981), The economics of superstars, American Economic Review, 71(5), 845–58.

Rosen, S. (1983), Specialization and human capital, Journal of Labor Economics, 1(1), 43–49.

Rysman, M. (2009), The economics of two-sided markets, Journal of Economic Perspectives, 23(3), 125–143.

Rysman, M. (2004), Competition between networks: A study of the market for yellow pages, Review of Economic Studies, 71, 483–512.

Saxenian, A.L. (1996), Inside-out: Regional networks and industrial adaptation in Silicon Valley and Route 128, Cityscape: A Journal of Policy Development and Research, 2(2), 41-60.

Schultz, N., F. Parisi and B. Depoorter (2002), Fragmentation in property: Towards a general model, Journal of Institutional and Theoretical Economics, 158, 594–613.

Scotchmer, S. (1991), Standing on the shoulders of giants: Cumulative research and the patent law, Journal of Economic Perspectives, 5(1), 29–41.

Shannon, C.E. (1948), A mathematical theory of communication, Bell System Technical Journal, 27, 379–423 & 623–656.

Shapiro, C. and H. R. Varian (1998), Versioning: The smart way to sell information, Harvard Business Review, 1998 (November-December), 106–114.

Shapiro, C. and H. R. Varian (1999), Information rules: A strategic guide to the network economy, Harvard Business School Press, Boston, MA.

Shy, O. (2001), The economics of network industries, Cambridge University Press, Cambridge.

Sokol, D. and R. Comersford (2016), Does antitrust have a role to play in regulating big data?, mimeo.

Stephan, P.E. (1996), The economics of science, Journal of Economic Literature, 34(3), 1199–1235.

Stiglitz, J.E. (2000), Contributions of the economics of information to twentieth century economics, Journal of Quarterly Economics, 115(4), 1441 –1478.

Stucke, M. and M. Grunes (2015), Debunking the myth over big data and antitrust, Antitrust Chronicle, 5.

Surblyte, G. (2016), Data as a digital resource, Max Planck Institute for Innovation and Competition Research Paper No. 16-22.

Tucker D. R. and H. B. Wellford (2014), Big mistakes regarding big data, Antitrust Source, American Bar Association.

Umbeck, J. (1981), Might makes right: A theory of the formation and initial distribution of property rights, Economic Inquiry, 19(1), 38–59.

Van Alstyne M., E. Brynjolfsson and S. Madnick (1995), Why not one database? Principles of data ownership, Decision Support Systems, 15, 267–284.

Varian, H., J. Farrell and C. Shapiro (2004), The economics of Information Technology: An Introduction. Cambridge: Cambridge University press.

Viard, B. (2007), Do switching costs make markets more or less competitive?: The case of 800-number portability, RAND Journal of Economics, 38(1), 146 – 163.

Vigo, R. (2013), Complexity over uncertainty in generalized representational information theory (GRIT): A structure-sensitive general theory of information, Information, 4, 1–30.

West, J. (2007), The economic realities of open standards: Black, white and many shades of gray. In: Greenstein, S. and V. Stango (eds.), Standards and public policy, Cambridge: Cambridge University Press, 87–122.

Wiebe, A. (2016a), Protection of industrial data – A new property right for the digital economy?, Journal of Intellectual Property Law & Practice, 12(1), 62–71.

Wiebe, A. (2016b) Who owns non-personal data? Legal aspects and challenges related to the creation of new 'industrial data rights', Slides presented at the GRUR conference on data ownership, Brussels.

Zech, H. (2016), Data as a tradeable commodity, in: De Franceschi, A. (ed.), New Features of European Contract Law: Towards a Digital Single Market, Cambridge, 51–79.

Zech, H. (2015), Information as property, JIPITEC, 6(3), 192–197.

Zhu, H., S. E. Madnick and M. Siegel (2008), An economic analysis of policies for the protection and reuse of non-copyrightable database contents, Journal of Management Information Systems, 25(1), 199–232.

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub