European Commission

JRC TECHNICAL REPORT

JRC Digital Economy Working Paper 2020-09

An economic perspective on data and platform market power

Martens, B.

2021

Joint Research Centre

Contact information
Name: Bertin Martens
Address: c/Inca Garcilaso 3, 41092, Seville (Spain)
Email: Bertin.MARTENS@ec.europa.eu
Tel.: (+34) 954.488.366

EU Science Hub
https://ec.europa.eu/jrc

JRC122896

Seville: European Commission, 2021.

How to cite this report: Martens, B., *An economic perspective on data and platform market power*, JRC Digital Economy Working Paper 2020-09, JRC122896.

Contents

Abstract

This paper starts with some basic economic characteristics of data that distinguish them from ordinary goods and services, including non-excludability and non-rivalry, economies of scope in data re-use and aggregation, the social value of data and their role in generating network effects. It explores how these characteristics contribute to the emergence of large digital platforms that generate a combination of positive and negative welfare effects for society, including data-driven network effects. It distinguishes between lexicographic and probabilistic data-driven matching in networks. Both may lead to market "tipping". It emphasizes the social value of data and the positive and negative social externalities that may come with this. Platforms are necessary intermediaries to generate the social welfare or network externalities from data. However, the economic role of data-driven platforms is ambivalent. On the one hand, platforms enable society to benefit from positive externalities in data collection via economies of scale and scope in data aggregation of transactions and interactions across users, both firms and consumers. That gives them a privileged market overview that none of the individual users has. Platforms can use this information asymmetry to facilitate interaction and increase welfare for users. These data externalities attract users to the platform. On the other hand, data-driven network effects may result in monopolistic market power of platforms which they can use for their own benefit, at the expense of users. Any policy intervention that seeks to address the market power of online platforms requires careful balancing between these two poles. Finally, the paper briefly discusses ecosystems that leverage data to coordinate interactions between different platforms.

1 Introduction

Data are the driving force behind the digital economy, including the large online platforms that have emerged as key players in the digital economy. Data are collected, analysed, transformed, accessed and traded between many players in the digital economy. This paper presents some basic economic characteristics of data that distinguish them from ordinary goods and services. It explores how these characteristics contribute to the emergence of large digital platforms that generate a combination of positive and negative welfare effects for society.

Data access and trade may cover a variety of modalities of data exchange between two or more parties, ranging from monetised trade in data, to voluntary free access or the exchange of data in return for a service. Any voluntary data exchange is a market-based data transaction. A key question for data policy makers is whether private and voluntary data access decisions maximize the social welfare of society as a whole. Economists define market failures as situations where the aggregate private welfare of firms and consumers remains below the total welfare that society as a whole could achieve with a given technology. This occurs when the incentives of private firms and/or consumers make them behave in ways that diminish overall social welfare. This may justify regulatory intervention in data markets and the imposition of remedies to address these failures. These may include some form of mandatory data access conditions that overrule private decisions.

In line with the European Commission's "Better Regulation Guidelines"[1] we follow a broader approach to possible regulatory intervention in data and data-driven services markets. It includes monopolistic market failures that are usually handled by competition law but extends to other sources of market failures such as externalities, asymmetric information and missing markets because of high transaction costs. We also point out potential regulatory failures and social concerns, such as welfare distribution and discrimination that could motivate regulatory intervention. In contrast with the mainstream competition law and economics view that adheres to a narrow consumer welfare policy objective[2], this paper takes a wider public policy economics view and focuses on the overall social welfare of society as a policy objective, combining welfare of firms and consumers. The distinction between consumer and social welfare may become important for example in data-driven online platforms. Policies that focus exclusively on the consumer side may have unintended negative effects on the supply side of the platform, and vice versa.

Furthermore, we go beyond markets and look at the impact that data have on institutions and organisational arrangements in the digital economy. A striking feature of this new organisational landscape is the emergence of online platforms and possibly platform ecosystems that create structural links between several platforms[3]. Digital data technology contributed to the emergence of new markets for goods and services that were not feasible in the pre-digital economy because the technology was simply not available to overcome some information cost constraints. These new markets often require new ways of organising economic exchange and new types of firms that are generically labelled as "platforms". At the same time, these platforms may generate new sources of market failures that will also be discussed.

This paper is structured as follows. Section 2 discusses the specific economic characteristics of data that are in several respects different from ordinary goods and services. We explore how these characteristics affect data collection – or the market for data exchange between data sources and collectors – and data trade between data collecting and data using firms. Section 3 brings platforms into the picture, a new type of firms that leverages the economic characteristics of networks and data to create more efficient markets. We examine the benefits that this brings as well as potential new sources of data-driven market failures. In Section 4 we move from single platforms to data-driven ecosystems where platforms coordinate their activities through data sharing. Section 5 adds some concluding observations.

---

1 The European Commission's "Better Regulation Guidelines" (2017) are available at https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how/better-regulation-guidelines-and-toolbox_en The European Commission's "Data Strategy" (2020, p 14 footnote 39) also advocates a market-failure based approach to regulatory intervention in data markets. Available at https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

2 Jason Furman, Diane Coyle, Amelia Fletcher, David McAuley and Peter Marsden, (2019), Unlocking Digital Competition, Report of the Digital Competition Expert Panel. Stigler Committee on Digital Platforms, (2019)

3 Michael G. Jacobides Carmelo Cennamo Annabelle Gawer (2018) Towards a theory of ecosystems, Strategic Management Journal, May 2018.

2 The economic characteristics of data

2.1 Data are usually an intermediary input, not a final consumer good

Unless they are aviation aficionados, consumers do not search for flight schedules on Google or Skyscanner because they enjoy looking at these schedules but because they want to buy an air transport service. Data are not created ex nihilo. They are collected from observations on the behaviour of people, machines and nature – the data originators. Firms collect data directly or from consumers or other firms. They can then be used for the production or improvement of a good or a service. Data exchanges thus involve at least two markets, an upstream data collection market and a downstream data use market. There can be a single vertically integrated single firm operating on both markets, or there can be different firms in both markets that trade data between them. Data collection can happen prior to their use in services, or it can be a by-product of services. For example, Google Search collects data by scanning webpages while the Search ranking depends on data collected from users of the Search engine. Data exchange, data trade, data sharing and access are labels that may cover different exchange modalities: data can be traded for a monetary compensation or in exchange for a service, sharing can be for free or subject to conditions in other markets, etc. Data can be traded directly – when they are effectively transmitted between parties - or indirectly – when parties do not transmit data but only a data-driven service. For example, online advertising platforms like Google do not transmit consumer data directly to advertisers. They sell a targeted advertising service based on consumer data which they keep in-house.

2.2 Data collection has an economic cost

The data collector needs to have a financial incentive to invest in data infrastructure, for example because it offers the prospect of monetizing the data. Data originators, consumers and firms, need incentives to share their data with a collecting firm. A frequently observed business model in data collection markets is to offer originators a free service in return for sharing their personal or industrial data. The willingness of data sources to share data with collectors will not only depend on conditions in the data market but also on subsequent use of the data in services markets. For example, the willingness of consumers to share their data with a website will depend on the quality of services offered by that website as well as subsequent use of the data by the website, for instance for online advertising. Lack of transparency in data re-use may of course blur that picture. Firms that offer free services need to find a way to cover the cost of providing these services. Google and Facebook offer users free services in return for the ability to monetise user data in an online targeted advertising market. Just like in the real economy, there are no free lunches in the data economy though the party that pays for the lunch may be different from the party that enjoys the lunch. Any change in the cost of data collection and in the benefits for data users will affect the volume and possibly the quality of data collected.

2.3 The value of data depends on their use

Data have no value on their own; they become valuable only to the extent that consumers and firms can use them to improve their position in data-driven services markets. Data can have many effects on services markets. Economists have tried to get a better understanding of these market effects and the welfare impact on stakeholders. There is no coherent framework yet for the economic analysis of data. Some authors[4] focus on the revenue-shifting potential of data. They assume that a "better" dataset generates more revenue for a firm, for a given level of utility provided to users. If firms extract more revenue than the utility they provide to users, users will shift to other firms, unless the firm's monopolistic market power prevents users from moving. That would result in an anti-competitive use of data. Pro-competitive[5] uses imply that both firm revenue and user utility from the data-driven services increases with additional data. For example, more data collection and more efficient use of the data in a hotel booking platform can simultaneously improve the user experience, revenue for hotels and platform revenue. Competitive use may still cause welfare shifts between

[4] De Cornière and Taylor (2020) "Data and competition: a general framework with applications to mergers, market structure and privacy policy". Mimeo, February 2020.

[5] The notions of pro- and anti-competitive behaviour go beyond classic notions of competition policy. They take a broader social welfare perspective that combines firm revenue and consumer utility. Increased market shares and market power can still be pro-competitive if they increase overall welfare.

firms and their customers, or between sub-groups on each side. This may trigger equity and welfare distribution concerns, for example when firms use data for price- or other forms of discrimination strategies that increase the welfare of the firm but not for some users. A problem with this very generic approach to data is that all these statements are subject to empirical evidence. This may be easy to obtain for firms and platforms that collect user data and run behavioural experiments with their online users in order to decide on their profit-maximizing commercial strategies. It is more difficult for policy makers to access relevant data that provide insights that could feed social welfare improving policies[6].

2.4 Excludability and monopolistic data trade

Contrary to physical goods, data are not excludable by nature. They can easily be copied and disseminated. The law can assign exclusive rights to data originators and/or collectors. So far, there are no general data ownership rights in the EU or elsewhere[7]. In a few cases, the law grants *erga omnes* exclusive rights. For example, in an attempt to bring data rights in line with the principles of intellectual property rights, the EU Database Directive[8] granted, under restrictive conditions, *sui generis* ownership rights to data collectors, the producers of databases. The EU General Data Protection Regulation[9] (GDPR) grants some exclusive and inalienable rights to natural persons as data originators, to keep control over their personal data, including the right to consent to access to personal data, and data access, portability, and deletion for the data subject. In the case of personal data, the data subject as data originator is usually unambiguously defined. This is not necessarily the case for non-personal machine-generated data that may be co-generated and collected by several parties. Assigning exclusive rights to any of these parties may affect the entire value chain in an industry[10]. All these attempts at assigning exclusive rights over data to private parties are in line with the Coase Theorem that hypothesizes that markets will work efficiently when ownership rights are well-defined and transaction costs are low or zero. However, because of the intrinsic social value of many data and the inability of individuals to internalize the externalities that their use entails, private ownership rights cannot bridge the gap between the private and social value of data (see Section 2.8 below).

In the absence of legal ownership protection, a data holder can apply technical protection measures to ensure his exclusive control and access to the data. This makes him a *de facto* data monopolist, provided there are no close substitute data sources. Exclusive access is necessary to raise revenue from selling data or data-driven services via negotiated bilateral contracts with data users that determine data access and use conditions, including prices. Contracts between contracting parties benefit from legal protection under commercial law. They can be enforced in courts. However, they cannot be enforced against third-parties. In case of data leaks, data holders have no recourse against third-parties that benefit from these leaks.

If more parties have access to the same dataset, or to close substitutes, competition will drive prices down to the marginal cost of reproduction, which is usually close to zero for digital data. That eliminates any opportunities to generate revenue from the data and any incentives to invest in the collection of data. Monopolistic data pricing above marginal cost requires rationing or reducing the quantity and/or quality of data that can be accessed. Not all demand will be satisfied, unless perfect price discrimination between buyers would be feasible. Monopolistic trade does not maximize social welfare[11]: it increases the welfare of the data holder at the expense of data users. Completely open data markets on the other hand drive prices and revenue down to zero and eliminate incentives to collect data in the first place. Data policy requires careful balancing between these two extremes.

6 This explains the origins of a range of Business-to-Government data sharing initiatives in several EU Member States and by the European Commission. See for example the report of the Expert Group on B2G data sharing, available at https://ec.europa.eu/digital-single-market/en/news/meetings-expert-group-business-government-data-sharing

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

8 Directive 96/9/EC of the European Parliament and the Council of the European Union, 11 March 1996, on the legal protection of databases.

9 Regulation (EU) 2016/679 of the European Parliament and of the Council of 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.

10 For an example from the agricultural sector, see Can Atik and Bertin Martens, Governing agricultural data: Comparing the US and EU Agricultural Data charters, JRC Digital Economy working paper, December 2020.

11 Unless there is perfect price discrimination in data sales. This is rarely feasible because it requires a lot of information on the willingness to pay of data buyers.

2.5 Data are not a homogenous product

They can be traded in various levels of fine graining and information content. They are subject to quality differentiation. For example, detailed consumer profile data are more valuable than coarse-grained data. Quality differentiation may be required in order to avoid falling into the Arrow Paradox: once data are revealed to a potential buyer there is no point in trading them anymore because the buyer already has the information he wanted to buy. There are many strategies that a potential data seller can apply to reduce the quality or information content of a dataset in order to overcome this paradox.[12]. He can offer a reduced sample of the data, or a coarse-grained or aggregated version that does not reveal details, or an anonymized version, etc. For example, mobile phone operators sell mobility insights but not original consumer data; data will be anonymized, aggregated and processed. The seller can also refrain from sharing data directly with a buyer and deliver an indirect data-based service only, as in the Google advertising example. Data sellers vary the quality of data collected from originators and transmitted to final users to maximize their profits[13]. Collecting very detailed data from consumers may make them suspicious and reject the service offered in return for the data. Handing over too detailed consumer data to data users may have a similar effect on the originators. Data buyers will want detailed data because it enables them to discriminate in the sale of their services. The data intermediary will adjust the quality of the data that he collects and sells (the level of aggregation and segmentation) to maximize his profits.

2.6 Non-rivalry and economies of scope in data re-use

Data are non-rivalrous. Many parties can use the same dataset at the same time for a variety of purposes without functional loss to the original data collector. Rival goods can only be used by one party at the time. For example, a car is a rival physical good and can only be used by one driver at the time. If a car would be non-rival, all drivers could re-use the same car at the same time to drive to different destinations. The welfare gains would be enormous: it would suffice to invest in the production of a single car to cater to the needs of all drivers. Data collected by one firm can be re-used for other purposes, either by the same firm or by other firms provided they can access the data. It results in cost savings because the primary data collection effort is a sunk cost that can be amortized across many uses, rather than remaining confined to a single user. It can boost innovation and enable the production of new and innovative data services that the original data collector had not envisaged. This promise of substantial welfare gains from exploiting non-rivalry in data re-use constitutes the foundation stone of the data access and sharing debates[14].

Economies of scope in re-use were originally defined in the context of joint production and (re-) use of the same product or asset to produce other outputs[15]. For example, a car manufacturer can re-use the same engines in different car models. Re-use of the same non-rival engine design entails zero marginal re-design costs. However, there is a positive marginal cost for physical re-production of additional engines. Non-rival immaterial products, such as knowledge and digital data, have quasi-zero marginal re-production costs because it involves only copying an electronic data file. Note that data re-use by other firms may create interoperability problems and important fixed costs for the design of an interface.

Data re-use and access by other parties also has a cost side. All digital data can, in principle, be made interoperable and shared for the benefit of society[16]. However, neither firms nor individuals want their private data to be widely available. Privacy and commercial confidentiality are important for the autonomy of private decision-making and for extracting private value from these decisions. While non-rival data can be shared by firms and individuals without functional losses, sharing may entail an opportunity cost and economic losses for the original data holder. Other firms may re-use the data in services applications that compete with those of the original data collecting firm and undermine the latter's market position[17]. The data holder may want to

---

12 For an overview, see Dirk Bergemann and Alessandro Bonatti, 2018, Markets for information: an introduction, CEPR discussion paper DP1314, 2018.

13 Dirk Bergemann, Alessandro Bonatti, and Tan Gan (2020), The economics of social data. Yale University, Cowles Foundation discussion paper nr 2203 revised.

14 OECD Maximizing the economic and social value of data, understanding the Benefits and Challenges of Enhanced Data Access, Directorate for Science and Technology, Committee on Digital Economic Policy, Paris, November 2016. Charles Jones and Christopher Tonetti, Nonrivalry and the economics of data, NBER Working Paper nr 26260, September 2019.

15 David Teece, Economies of scope and the scope of the enterprise, Journal of economic behaviour and organisation, 1980. David Teece (1982) Towards an economic theory of the multi-product firm, Journal of economic behaviour and organisation, 1982, pp 39-63. John C. Panzar and Robert D. Willig, Economies of Scope, the American Economic Review, Vol. 71, No. 2, May 1981.

16 John Palfrey and Urs Gasser, Interop: The Promise and Perils of Highly Interconnected Systems, 2012.

17 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.

produce these alternative services in-house and appropriate the benefits. Firms may re-use personal data for purposes that harm the data subject's privacy and welfare.

Firms and persons will trade off the expected benefits from data sharing against the expected costs and risks that they might incur from doing so. These private cost-benefit perceptions may limit the extent of data exchange, sharing and re-use. The question for policy makers is whether private data decisions by consumers and firms maximize the welfare that society as a whole could derive from the data. If not, there is a market failure that may require policy intervention. Policy intervention should not seek to maximize data sharing. Data sharing is not an objective in its own right but a means to achieve higher social welfare for society. Policy makers should only intervene when the market is not delivering a social welfare-maximizing volume of data sharing, considering both the costs and benefits of data sharing.

2.7 Economies of scope in data aggregation

A second, and often neglected, source of economies of scope in data comes from data aggregation. Merging two complementary datasets can generate more insights and economic value compared to keeping them in separate data silos, provided that the datasets are complementary and not entirely separable. This insight can be traced back to the economics of learning and division of labour. Rosen[18] observed that when a person has a choice between learning two skills, specialisation in one skill is always beneficial when the costs of learning both skills are entirely separable. However, when learning costs are not separable because knowledge sets are complementary, there are economies of scope in learning both skills, provided the benefits from combining the two exceed the additional learning costs. This insight can be applied to data. When two datasets are complementary, applying data analytics – the equivalent of learning – to the merged set will yield more insights and be more productive than applying it to each set separately, especially when the marginal cost of applying analytics to a more complex dataset is relatively small.

Economies of scope in data are controversial in economics, also because they are often misunderstood. Authors usually do not distinguish between economies of scale and scope, or between economies of scope in re-use and in aggregation of data. Tucker[19] defines economies of scope somewhat ambiguously as cost savings relative to an "increased level of production of multiple products". "Increased level of production" refers to economies of scale. "Multiple products" could be interpreted as economies of scope in re-use of data but not in data aggregation. A useful way to distinguish economies of scale and scope is to consider a dataset as a two-dimensional spreadsheet, with the number of columns representing the number of variables and the number of rows the number of observations on these variables. Economies of scale refer to increased prediction accuracy due to an increase in the number of rows. Economies of scope in data aggregation refer to increased prediction accuracy due to an increase in the number of columns. Adding more columns (variables) is not helpful when they are highly correlated or when they are not related at all. A number of empirical studies claim that economies of scope in data are weak or non-existent[20]. All these studies are more about economies of scale rather than scope. Bajari et al[21] come closest to a proper study of economies of scope in aggregation when it merges data across several product markets. They find that product sales forecasts do not become more accurate when historical data from several products are combined. However, weak complementarity among product markets results in highly separable datasets and thus in weak economies of scope. The absence of empirical studies on economies of scope in data aggregation is a major gap in data economics. There is anecdotal evidence in support of economies of scope in data aggregation. McNamee[22] explains how Google gradually improved its targeted advertising by combining personal data from several sources, starting from web searches and adding email and maps (location) data. Navigation apps like Waze and Tom-Tom combine real time GPS location data with maps that are populated with data from a wide range of public and private sources including road and traffic authorities, municipalities, firms and in-map advertisers. These public sector data may have little commercial value on their own but create a valuable service when aggregated with other data.

[18] Sherwin Rosen (1983) Specialisation and human capital, Journal of Labor Economics, Volume 1, Number 1 Jan., 1983.

[19] Tucker (2019, p 5)

[20] Chiou and Tucker find no decrease in search engine accuracy when time series of consumers' historical searches are shortened because of EU privacy regulation. Neumann et al show that large data brokers do not necessarily perform better in consumer profiling than data brokers with fewer consumer profile data. Claussen et al find that more individual user data helps algorithms to outperform human news editors but decreasing returns to user engagement set in rapidly. Schaefer et al find that the quality of search results improve with more data on previous searches. McAfee et al find that Google Search outperforms Microsoft Bing in long-tail searches because of a higher number of users.

[21] Bajari et al (2018)

[22] Roger McNamee (2019) Zucked, waking up to the Facebook catastrophe.

Economies of scale and scope in data aggregation are a source of positive externalities. In the age of artificial intelligence and machine learning, personal data collected on the behaviour of one set of consumers has predictive value for the behaviour of other consumers[23]. Once a firm has accumulated a critical mass of consumer data, the additional insights obtained from adding another consumer's personal data are small[24], compared to what can be learned from data already collected about persons with a similar profile. Acemoglu et al[25] argue that these externalities in personal data collection create a market failure. They diminish the value of individual personal data as well as consumer incentives to protect their privacy. That, in turn, increases the supply and further decreases the market value of personal data. Data collectors can reap the benefits of that externality; consumers cannot prevent this negative externality for their own data. Their best deal is to exchange their personal data in return for a free online service that has a higher marginal use value for them than the depressed market value of their individual data. This externality could explain the privacy paradox[26]. Consumers value their privacy but do not invest in protecting it. They have relatively little resistance to sharing location or browsing data when used for advertising purposes. Consumer resistance is high only for the most sensitive data such as bank statements or fingerprints[27]. Investment in privacy protection tools may have a signal value in itself that can be exploited against consumer interests[28].

Note that the two interpretations of economies of scope in data (re-use and aggregation) may lead to very different policy implications. Economies of scope in re-use provide an argument in favour of data dissemination and de-concentration. Economies of scope in aggregation, by contrast, favours data concentration in large pools from a variety of sources. The two are not mutually exclusive. Since data are non-rival they can be stored at the same time in concentrated pools and in distributed settings. Both concentration and de-concentration can result in market failures that undermine social welfare[29].

2.8 The social value of data

A peculiar characteristic of many[30] data is their social use value. Economies of scope in aggregation adds a social dimension to the value of data. Owners of two separate but complementary datasets can reach a higher level of value and insights from their data if they pool the two sets. Another source of social value of data is related to economies of scale. Once a sufficiently large sample of behavioural observations has been compiled to produce robust predictions, that data sample can be used to predict the behaviour of agents outside the sample[31]. This implies that collecting more data about other agents with similar characteristics has diminishing marginal value because the existing dataset is sufficiently representative to predict the behaviour of other agents, even if other agents refuse to share their personal data.

These externalities imply an inherent market failure in exclusive private control over complementary data, both for data sources and for data collectors. The party that does (not) provide the data to a collector is not necessarily (may still be) the party that is affected by their use. The *de facto* exclusive data holder is not necessarily the party that maximizes benefits from the data. Two or more parties can agree to pool their data and generate the full social value of the data. However, coordination costs and risks may undermine this spontaneous pooling. An intermediary agent may be required to realize the social externalities from data pooling and turn them into benefits that (a) pay for the coordination costs, (b) generate benefits that incentivise individuals to participate in the pool, and (c) extract a profit from the intermediation services. With this, we reach the world of data platforms in the next section.

---

23 Ajay Agrawal, Avi Goldfarb, and Joshua Gans (2018), Prediction Machines: The Simple Economics of Artificial Intelligence, Harvard Business Review Press, 2018.

24 For an example of diminishing returns to scale in data collection, see for example Maximillian Schaefer and Geza Sapi (2019) Data Network Effects: The Example of Internet Search, mimeo, December 2019.

25 Acemoglu et al (2019)

26 Acquisti et al, (2016).

27 Jeffrey Prince and Scott Wallsten (2020) How Much is Privacy Worth Around the World and Across Platforms?, Technology Policy Institute.

28 Dengler and Prüfer, 2018.

29 Economies of scope in aggregation and re-use exist in other domains too, for example in intellectual property rights. For example, the market value of a set of complementary patents may be higher than the sum of their separate values. Hence the practice of patent bundling and thickets, and the bundling of Standard Essential Patents (SEPs) to facilitate re-use of technical standards. Bundling strengthens the monopolistic position of patent holders. Fair, reasonable and non-discriminatory (FRAND) licensing seeks to compensate this by avoiding abusive behaviour.

30 Some types of data may have little or no social value because they remain situation, person or firm-specific and cannot be used by other agents or situations, or has no complementarity with other datasets.

31 Dirk Bergemann, Alessandro Bonatti, and Tan Gan (2020) The economics of social data, March 2020, Cowles foundation discussion paper nr 2203R. Acemoglu et al (2019).

3 Platforms and data-driven network effects

Much of the current debate on data access is still implicitly set in the context of traditional firms and data exchanges between individual data collectors and re-users. However a substantial volume of data exchanges and data-driven services trade takes place in a new type of firms that are usually classified under the generic label of "platforms". While monopolistic market failures may occur in linear data exchanges, recent data and competition policy related reports[32] pointed out that monopolistic behaviour occurs mainly in very large online platforms that have become gatekeepers to online markets. In this section we explore the crucial role of platforms in the digital economy and the role that data play in these platforms.

3.1 Platforms in the digital economy

What are platforms? There are many definitions of platforms – or multi-sided markets in economic jargon – in the economic literature and there is no consensus among economists on these definitions[33]. The first generation of multi-sided market models were extension of the economics of infrastructure networks, such as telephone and railroad networks for example. Network effects or network externalities occur when users derive benefits from the presence of other users. When more users connect to a telephone network this creates more opportunities to call other users. It makes the network more attractive to all users. The first generation of platform models in economics[34] focused on markets with at least two types of users, for instance buyers and sellers. Platforms are faced with a "chicken and egg" problem: they need many users on one side of the market in order to attract many users on the other side of the market. They can solve this problem by charging a very low or zero price to one side of the market to attract many users on that side, and charging a high price to the other side to pay for the cost of operating the platform. Users on the side with a high price elasticity of demand pay low or zero entry costs while users with low price elasticity of demand pay a higher price. This explains why advertisers pay for ads while users get free access to search and social media services: advertisers have no choice but to advertise in a particular platform where a user with a specific profile is looking for a good or service that the advertiser can offer. Users can, to different extent, however multi-home between many platforms to find what they are looking for. These models ran into problems to distinguish between intermediary platform and ordinary retailers and defining the type of interaction between two sides[35].

Recent economic thinking on platforms has broadened the definition. Platforms can be defined[36] more generically as undertakings that bring together economic agents and actively manage network externalities[37] between them. The key role of platforms is to generate positive network effects or network externalities and in this way maximize the social value that can be extracted from the data collected by the platform. The presence of economies of scale and scope in the data aggregated by a platform ensure that the collective social value of data exceeds the sum of their individual private values[38]. Creating a searchable catalogue of products or a directory of users is a first step in generating that social value. For more efficient probabilistic matching, the platform requires detailed data on buyer characteristics and preferences and on the

[32] Crémer, J., Y-A de Montjoye and H. Schweitzer (2019) Competition policy for the digital era, Report for Commissioner Vestager, European Commission. Furman, J., D. Coyle, A. Fletcher, D. McAuley and P. Marsden (2019), Unlocking digital competition, Report of the digital competition expert panel, March 2019. Scott-Morton, F., P. Bouvier, A. Ezrachi, B. Julien, R. Katz, G. Kimmelman, A.D. Melamed, and J. Morgenstern (2019), Report of the Committee for the study of digital platforms, market structure and antitrust subcommittee, George Stigler Center for the study of the economy and the state, University of Chicago Booth School of Business, May 2019.

[33] For an overview of the (fairly recent) history of economic thinking on platforms, see for example Bertin Martens, 2016. "An Economic Policy Perspective on Online Platforms," JRC Working Papers on Digital Economy 2016-05, Joint Research Centre.

[34] Caillaud, B., & Jullien, B. (2003). Chicken & egg: Competition among intermediation service providers, The RAND Journal of Economics, Vol. 34, No. 2 (Summer, 2003), pp. 309-328; Parker, G & Alstyne, M.. (2005). Two-Sided Network Effects: A Theory of Information Product Design. Management Science, 51(10), 1494-1504. Rochet, J.-C., & Tirole, J. (2003). Platform Competition in Two-sided Markets. Journal of the European Economic Association, 1(4), 990-1029. Rochet, J.-C., & Tirole, J. (2006). Two-sided markets: a progress report. Rand Journal of Economics, 37(3), 645-667.

[35] Hagiu, A., & Wright, J. (2015). Marketplace or Reseller? Management Science, 61(1).

[36] Jens-Uwe Franck and Martin Peitz (2019), Market definition and market power in the platform economy, CERRE, Brussels, May 2019. This definition does avoids the problem of setting a minimum number of market sides; one is enough.

[37] Despite the centrality of network effect in these economic models, there is surprisingly little empirical research that estimates the strength of these network effects. Existing empirical evidence is based on offline physical networks. There is hardly anything on online platforms. (references?)

[38] Bergemann, Bonatti and Gan (2020) The economics of social data. Yale university, Cowles Foundation discussion paper nr 2203 revised.

characteristics of the products and services offered. For example, Netflix can improve its film title search engine when it learns more about user preferences and film characteristics[39].

A comparison with traditional offline markets illustrates the importance of the online platform's role as data collector and producer of data-driven externalities. In a traditional town market buyers walk around between market stalls, collect information on what is on sale and sales conditions, and make their choices. The town authority as market organiser has hardly any information on sellers' offers, buyer preferences and actual transactions. Each user has to collect this information separately; there is no common information pool. This is privately costly for users and socially costly for society as a whole. Costs increase with market size. In online markets, the platform operator collects an aggregated view of supply and demand and actual transactions. Users can benefit from this aggregated information. It would be impossible for users in large online platforms with millions of product entries to collect all information on their own. Platforms are in a unique position as third-party data aggregator to realize economies of scale and scope in data aggregation across many users. Individual users cannot realize these benefits. This fits well with our definition of economies of scope in data aggregation: the value of the insights from the aggregated dataset is higher than the sum of values of individual user datasets[40].

The label "platform" refers to a multi-sided market as well as to the firm that manages this market. While markets can grow spontaneously, in many cases they require an organiser to take the initiative and define the operating conditions. Platforms are new types of market-organising firms that emerged in the wake of digital data. The traditional view of the firm goes back to Ronald Coase[41]. Coase wondered why firms exist as an arrangement between workers who divide tasks and exchange intermediate goods and services in the organisational setting of the firm rather than going through the market for each of these exchanges. He argued that contractual arrangements reduce transaction costs compared to going through the market for each exchange between workers. The borderline of the firm, between in-house production and external trade, depends on transaction costs. Digital data and online platforms have dramatically reduced these transaction costs to quasi-zero in many cases. With quasi-zero digital transaction costs, some firms stop in-house production altogether, delegate production to external agents and transform themselves into market places. These firms "invert" and become market organisers rather than production organisers[42]. In contrast to traditional firms that keep the market outside, they organise a platform market where different types of users, for instance buyers and sellers, can trade goods and services. Iansiti and Lakhani[43] show that data-driven platforms are not subject to diminishing returns to scale. Human labour is replaced by data-driven algorithmic procedures with high fixed set-up costs but nearly zero marginal costs. Non-rival data and algorithms make these platforms infinitely scalable. This leads to huge productivity and efficiency gains but also to increased market power and monopolisation.

3.2 The role of data in platforms

The above explanations show that data collection and analytics play a key role in the intermediation function of platforms. However, the first generation of economic models of platforms actually had no explicit role for data. These models are suitable for relatively simple networks with unambiguous lexicographic matching between users, such as telephone networks that inspired these models. The data-free platform model fails to explain what happens in complex platforms where collecting user data is indispensable to generate data-driven network effects and increase matching efficiency in ambiguous and probabilistic matching[44].

Data play a role in generating network effects. In some cases the role of data is very minimal and static. For example, users in a telephone network differ only by their telephone number, a unique lexicographic address. Users can be unambiguously matched by combining two lexicographic addresses. The only dataset required to make the telephone network operate optimally is a telephone directory. Matching between telephone users cannot be improved by observing the behaviour of the users. Similarly, in simple online e-commerce stores, a

39 Marco Iansati and Karim Lakhani (2020) Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World, Harvard Business Press, chapter 6.

40 For example, Reimers and Walfogel (2020) estimate that the welfare effect of aggregated consumer book review data on the Amazon book sales platform is about 15 times larger than the welfare effects from a single-authored book review in a newspaper. This informational advantage puts platforms in a strong bargaining position vis à vis individual user data.

41 Ronald Coase (1937) The nature of the firm, Economica, November 1937.

42 See Parker, Geoffrey and Van Alstyne, Marshall W. and Jiang, Xiaoyue, Platform Ecosystems: How Developers Invert the Firm (August 17, 2016). Boston University Questrom School of Business Research Paper.

43 Marco Iansiti and Lakhani (2020)

44 Prüfer and Schottmuller, Op. cit.

targeted search for a well-defined product may just require a catalogue of unambiguously defined products. For example, search for a book title in the Amazon book store. In these cases network effects are driven by the numbers of products and users and their unique identification. The quantity and quality of data on users and products plays no role in unambiguous matching processes.

In other platforms the role of data is crucial. For example, matching in search engines and targeted advertising markets requires data on the characteristics of users and products, beyond a lexicographic identifier, in order to select the most likely and optimal matches. Many matches are possible but which match or ranking of matches is the most optimal? Probabilistic matching requires more detailed data on relevant user characteristics in order to improve the efficiency of matching. For example, a search engine will not only index the IP addresses of webpages in the world-wide web but also collect the content of the pages, analyse and classify that content. It will collect data on user clicks on search ranking in order to better understand which pages are most relevant for a specific search term and for a specific user. It will then carry out a probabilistic matching between the two, resulting in a ranking from most likely to less likely matches. More precise data on user preferences will increase the efficiency of probabilistic matching. The quantity, quality and analytics of the data will play an essential and dynamic role in generating data-driven network effects.

Many of today's largest online platforms are probabilistic matching services: Google Search, Facebook social media, online advertising, Amazon, Netflix, Uber, e-scooter platforms, etc. They put data at the core of their business model and specialise in transactions that require substantial datasets to do an efficient matching between users. Platforms help to create new markets that were missing in the pre-digital economy because information-related transaction costs were too high. For example, finding a hotel was costly in the analogue economy and required intermediation from travel agencies that offered a limited choice to consumers. Finding "information" in general was costly. These missing information markets were not a market failure because the technology to overcome them was not available at the time, or remained very imperfect. Digital data technology has dramatically reduced information cost and thereby expanded user choices.

Data-driven network effects[45] are intrinsically linked to economies of scale and scope in data aggregation. They can reinforce the efficiency of probabilistic matching networks and thereby strengthen network effects. For example, McAfee et al[46] show how larger number of users in Google Search make it more efficient in rare search terms compared to Microsoft Bing that has a much smaller number of users. Economies of scale means more observations on similar search term while economies of scope in aggregation imply collecting search results from a wider variety of search terms. Algorithms reinforce the value of the data through a feedback loop that builds on better predictions and learning-by-doing that, in turn, strengthens data-driven network effects. That difference in efficiency, in turn, motivates users to shift to Google. The two may reinforce each other. The rise of artificial intelligence and machine learning has further amplified economies of scale and scope in data aggregation. Machine learning is a very data-intensive technology. While human learners can learn a behavioural response from a few observations, machine learning algorithms often require huge numbers of observations to correctly learn an appropriate response.

Individual consumers or firms cannot achieve these data-driven network effects on their own. They require third-party intermediaries to collect, classify and analyse data in order to make efficient use of it. That is the role of platforms. Platforms are in a unique position to aggregate data on transactions and interactions across many users, including firms and consumers. They can realize the economies of scale and scope in data aggregation that drive the social value of data. If individual firms and consumers would keep their own transaction data, without sharing them with the platform, the collective social value of data would not be realized. That gives them a privileged comprehensive market overview that none of the individual users has.

Platforms can use this privileged data position in two ways. First, they can use it to produce search and matching services for users: helping consumers to find what they are looking for, helping businesses to find their customers, helping advertisers to better target their ads, etc. That generates positive welfare effects or network externalities that attract more users to the platform. Second, they can monetise these services for their own benefit. They charge platform entry fees to users, based on the insights from the classic platform model: the least flexible users pay the highest market entry price. Market entry prices transfer some of the user surplus to the platform operator. Platforms need to strike a balance between the pursuit of their own

[45] Data-driven network effects were first analysed by Jens Prüfer and Christophe Schotmüller, 2018 Competing with Big Data, TILEC Discussion Paper No. 2017-006. Stucke and Grunes (2016) "Big data and competition policy" already speculated that there is a link between economies of scope and network effects or network externalities. Tucker (op. cit.) is not convinced by this conflation of concepts.

[46] Op. cit.

profits and providing welfare benefits to users[47]. Going too far in one or the other direction undermines the viability of the platform and the social welfare benefits that it can produce. For example, not-for-profit intermediary platforms exist. But their financial capacity to invest in innovative data collection, analytics and service production for users will be curtailed. Conversely, for-profit firms that do not generate social welfare externalities also exist.

3.3 Market failures in platforms

Platforms can achieve monopolistic market power in several ways. In the traditional platform economics model, data-free network externalities may confer market power on platforms[48]. "Market tipping" occurs because users prefer to congregate on the largest platform. This strengthens the position of the incumbent platform at the expense of potential new entrants into the market who have to overcome the hurdle of network effects in order to successfully compete with the incumbent. The dominant platform becomes a gatekeeper to the market. It sets market conditions. The platform's exclusive access to aggregated user data reinforces its monopolistic market position[49]. The downside of the positive network externalities of platforms is exploitation of this monopolistic position to the detriment of users and actual or potential competing platforms. They may steer users to transactions that are more beneficial for the platform and less so for users. They may use the data to compete with their own business users and to foreclose aftermarkets or leverage their market position in adjacent markets. Monopolistic power concentration tendencies are inherent in the platform and data economy. However, an exclusive focus on reducing monopolistic market economic power may undermine the positive social welfare externalities from data aggregation. Policies should seek to maintain the benefits of data aggregation externalities while addressing the anti-competitive use of platforms' data advantages.

The strength of data-driven network effects plays a key role in tipping[50] and varies by type of platforms and the relevant data in these platforms. For example, in ride hailing and e-mobility platforms, network effects are very local. The platform may be organised on a global basis but network effects depend on local supply and users in cities. Expanding the supply in city A has no benefits for users located in city B, unless they happen to travel frequently between the two cities. This makes it easier for smaller local platforms to compete in local markets with global platforms. Hotel booking platforms are global however. Users search for hotels in many cities and platforms have to ensure a wide geographical variety of offers. This makes competition more difficult. Platforms can pursue deliberate strategies to tip the market in their favour, for example by increasing the costs of multi-homing or switching to other platforms. For example, drivers can easily switch between ride hailing platforms with little costs. To discourage drivers from switching, platforms may offer them an uninterrupted sequence of rides, with advance notice of the next ride before the on-going ride is completed. Platforms can try to differentiate their products from competitors' by adding innovative features. If these features can easily be copied by competitors they offer a less sustainable advantage.

Hagiu and Wright[51] illustrate how the value that platforms can extract from data is conditional on several factors. Improving the quality of insights and the matching efficiency of data can be subject to economies of scale. In some cases a few observations are sufficient to make an accurate prediction, in other cases diminishing returns to scale remain far away. For example, automated driving algorithms are still far from perfect despite millions of miles of accumulated driving data by leading firms such as Google/Waymo. The value of these insights depends on the market size. This is often true for artificial intelligence based applications in platforms that depend on large numbers of observations. Insights that benefit from externalities and can be extrapolated to a wide number of users have high value. For example, personalised music recommendations in Pandora, based on cumulative learning from individual users, cannot easily be applied to other users. Spotify's shared music recommendations by contrast benefit from strong network externalities because they are useful to many users. The accuracy of Google Maps for traffic predictions is subject to data-driven network effects because it increases with the volume of data collected from Android users.

---

47 Babur de los Santos and Sergei Koulayev (2017) Optimizing Click-through in Online Rankings with Endogenous Search Refinement, Marketing Science, 36(4): 542–564 2017.

48 Caillaud and Julien, 2003; Rochet and Tirole, 2006; Parker and Van Alstyne, 2005, op.cit.

49 Orla Lynskey, Grappling with "Data Power": Normative Nudges from Data Protection and Privacy, 20 Theoretical Inquiries L. 189 (2019), London School of Economics.

50 See Marco Iansati and Lakhani (2020), op.cit., chapter 6.

51 Andrei Hagiu and Julian Wright (2020), When data creates competitive advantage, Harvard business review, January 2020.

Several competition policy reports investigate the link between data and platform market power[52] or monopolistic market failure. They suggest some re-thinking of competition policy tools to take into account the specific nature of platforms as multi-sided markets and the complexity of data collection, analytics and use in data-driven platforms. This includes a revision of the relevant market doctrine and theories of harm, new measures of market power and dominance thresholds in multi-sided markets, accelerate competition procedures to stay ahead of market tipping, etc. Since data-driven network effects are often the cause of competition problems, these reports pay attention to data policy tools as a means to attenuate data-driven monopolistic behaviour, for example by opening access to exclusive datasets, or a variety of data pooling and data sharing modalities. Data access or sharing may prevent an upstream monopolistic data collector from foreclosing downstream services markets. For example, car manufacturers design the car data architecture to retain exclusive access to car data, which they can leverage to increase their share in aftersales services markets. Mandatory data access for other aftersales service providers can prevent this competition problem to occur[53]. Opening data access may backfire however. It may reduce rather than increase competition when the data are hoovered up by large platforms that can offer users additional advantages, based on economies of scope in re-use and aggregation with other data sources. For example, payment services offered by large platforms such as Apple and Google, or payment services on the WeChat social media app in China and perhaps in future on Facebook, may compete with payment services offered by local banks or smaller payment services start-ups. Google Android and Apple iOS are increasingly present in cars and may offer aftermarket services that compete with manufacturers. Since data are not a homogeneous product (see section 2.6.), data access and sharing can be fine-tuned to a degree of coarseness that preserves some incentives and advantages for the original data collector while still broadening competition in the market for data-driven services. That would require a careful balancing act and constant market and technology monitoring by regulators.

Data sharing obligations may have potential disincentive effects on data collection efforts. Data sharing with potential competitors will erode a platform's data aggregation monopoly, lower the value of the data and undermine the ability to monetise the data. In a multi-sided market, modifying access conditions on one side of the market will have implications for other sides. For example, forcing a search or social media platform to share consumer data with competitors may not only affect consumer privacy. It lowers entry costs into advertising and will force platforms to increase entry costs on the consumer side, or integrate new money-raising sides into the platform, in order to make up for the lost revenue.

Platforms are both a blessing and a curse in the data economy. They are necessary intermediaries to generate benefits from data aggregation, realize data-driven positive network externalities and thereby enable the emergence of new markets that were not feasible prior to the arrival of digital data. At the same time, exclusive control over the data allows gatekeepers to control the ecosystem and generate significant value for their intermediation services. They can impose excessive entry and access conditions, and exclusive dealing rules preventing sellers from promoting their offers outside the gatekeeper's platform. Refusal to share the data with business users in the platform, or with competing platforms, gives them a competitive advantage that gatekeepers can use to foreclose the market and strengthen their monopolistic position, to the detriment of user welfare.

The European Commission's "Better Regulation Guidelines" distinguish between several types of market failures that may require regulatory intervention to ensure optimal production of social welfare for society as a whole. Besides monopolistic market failures, other sources of failure include externalities, information asymmetries and missing markets because of high transaction costs and risks. Besides market failures, regulators may also intervene in case of social concerns such as discrimination and unequal distribution of welfare. In the next sections we discuss three types of data-driven non-monopolistic market failures: negative externalities from data aggregation, asymmetric information problems that distort decision making by data users, and newly missing markets that emerge in the wake of the data economy because of high data transaction costs and new sources of data-related risks.

52 Crémer et al, op.cit., Furman et al, op.cit., Scott-Morton et al. op.cit.

53 Martens B. and F. Müller-Langer (2020) Access to digital car data and competition in aftersales maintenance markets, Journal of Competition Law and Economics, 2020. Wolfgang Kerber (2019) Data Governance in Connected Cars: The Problem of Access to In-Vehicle Data, Journal of intellectual property and information technology, vol 9(3) 2018. A similar argument can be made for giving payment service providers access to bank account data, as in the EU Second Payment Services Directive (PSD2), Directive (EU) 2015/2366.

3.4 Negative information externalities

So far we discussed the role that platforms play in generating data-driven positive network externalities. In this section we turn to negative data-driven externalities. We present examples of negative externalities on the consumer side in personal data markets, and on the producer side in commercial data markets. Positive externalities increase social welfare, provided they can be captured and turned into economic value for users. Negative externalities are to be avoided because they reduce the welfare of user, or they should be internalised by the party that causes the externality.

A first example of a negative data externality is the impact of consumer platforms on the value of personal data. Data collected on the behaviour of one set of users has predictive value for the behaviour of other users[54]. Acemoglu et al[55] show that the marginal value of an individual's personal data is diminished by negative externalities from economies of scale in data aggregation in platforms and economies of scope in re-use of personal data. Once a firm has accumulated a critical mass of consumer data, the marginal return in terms of improved insights and additional value in the secondary re-use market – for example for advertising purposes – from adding another consumer's personal data are close to zero, compared to what can be learned from extrapolation from data already collected about persons with a similar profile. This reduces the marginal value of a single person's dataset. It also reduces incentives for consumers to protect their privacy since their profile can be assembled from data collected from other persons. That, in turn, creates an excessive supply of personal data. Consumers may invest in privacy protection. That in itself may have signal value that can be exploited against consumer interests[56]. Following the entry into force of the EU GDPR, an empirical study on the use of personal data for advertising in the travel industry[57] finds that 12 percent of consumers refuse consent to collect their personal data. However, the study also find that the reduction in the supply of available data increases the value of the remaining advertising data and, because of externalities, does not negatively affect the predictability of consumer responses to advertising. Another study shows that consumers underestimate the negative impact of sharing their personal data with a platform[58]. Data sharing improves matching efficiency and makes it easier for consumers to find what they are looking for. At the same time, the increased matching efficiency enables the platform to charge sellers a higher entry costs. That, in turn, pushes up consumer prices for products sold on the platform. Consumers are not aware of this second-round effect of data sharing.

Is this negative externality a market failure that requires regulatory intervention to be corrected? Individuals have no better alternative option to realize a higher value for their personal data. Brynjolfsson et al[59] present empirical evidence that "free" services platforms compensate the negative externality and, in fact, generate a large consumer surplus. Consumers thus trade personal data at nearly zero value for valuable online services at a zero "free" price. That suggests that the positive network externalities produced by platforms outstrip the negative externality on personal data. Consumers get more value out of it than they put into it. Charging positive prices to advertisers and negative prices to consumers for these externalities creates a price distortion, at least from a traditional economics perspective. Zero prices are often seen as a market distortion[60]. Trying to correct any of these market distortions may reduce overall social welfare because it would reduce the number of consumers, the volume of data and make the platform less interesting for advertisers and for other consumers. It creates a lose-lose-lose situation. Public opinion often goes in the other direction, as the quip "if you are not paying you are the product" suggests. Other authors go a step further and suggest that consumers should be paid for the "data work" that they contribute to platforms[61].

Negative externalities may also occur on the firm or supply side of platforms. Suppliers sell their goods and services through online platforms like Amazon, eBay. Platform operators collect and aggregate data on product characteristics, sales and consumer choices across many users. Once sufficient data are collected they can predict market responses to changes in product characteristics and prices. Platforms can use these

---

54 Bergemann, Bonatti and Gan (2020), op.cit.

55 Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian and Asuman Ozdaglar, Too much data: prices and inefficiencies in data markets; NBER Working Paper 26296, September 2019.

56 Dengler, Sebastian and Prufer, Jens, Consumers' Privacy Choices in the Era of Big Data (April 9, 2018). TILEC Discussion Paper No. 2018-014. Available at SSRN: https://ssrn.com/abstract=3159028

57 Aridor, Guy and Che, Yeon-Koo and Salz, Tobias, The Economic Consequences of Data Privacy Regulation: Empirical Evidence from GDPR (January 29, 2020). Available at SSRN: https://ssrn.com/abstract=3522845

58 Rishabh Kirpalani and Thomas Philippon (2020) Data sharing and market power with two-sided platforms, NBER Working Paper 28023.

59 Erik Brynjolfsson, Avinash Collis, W. Erwin Diewert, Felix Eggers, Kevin J. Fox (2019) GDP-B: accounting for the value of new and free goods in the digital economy. NBER Working Paper 25695, March 2019.

60 Gans, Joshua, The Specialness of Zero, Jan 2020, available at SSRN: https://ssrn.com/abstract=3486964

61 Posner and Weyl, 2019 Radical markets.

data to spot opportunities to enter the market with their own products and services and compete directly with independent suppliers on the platform[62]. They can also use the data bias search rankings inside the platform in favour of their own products. This is a form of data-driven foreclosure or self-preferencing that distorts competition. Data collected from suppliers and transactions are leveraged in favour of the platform.

3.5 Negative effects from asymmetric information

Asymmetric information between individual users and data-collecting platforms is an almost natural state in a data-abundant digital world. Platforms as data aggregators will always have more and better information on the markets that they cover, compared to individual platform users (persons and firms). It is not only a question of amount of information however. Platforms will manipulate the level of fine-graining of information that they collect from data originators. The willingness to share information with the platform depends on the level of detail and the use of the data[63]. Conversely, platforms may degrade the level of detail and introduce segmentation on the data user side of the platform in order to maximize profits from their exclusive data. As private profit-maximizing firms, platforms will use this information asymmetry to their advantage, trying to extract maximum revenue from this data intermediation role. Users may take sub-optimal decisions because of imperfect information signals received from platforms.

Platforms may also use the information to promote its own services and products, competing with service producers on the platform. For example, in July 2019 the European Commission opened an investigation into Amazon[64]. Amazon combines the roles of online retailer on its own account and market place for independent sellers. The platform may have used non-public data that it collects and generates about the activities of independent sellers to compete with its own sellers.

Researchers have observed that Amazon allegedly degraded the quality of information signals to consumer search results to favour Amazon sales and reduce the prominence of sales by independent sellers ("self-preferencing")[65]. The market-distorting effects of asymmetric information in favour of the platform operator is well-documented in empirical studies on all kinds of search engines[66]. Platforms apply business models that may be based on sales margins (for retailers), commissions on sales (for market places) or advertising revenue (pure information matchmakers). The incentives embedded in the business models affect search rankings and drive a wedge between user preferences and the financial interests of platform. For example, hotel booking platforms can manipulate search rankings towards price offers that increase their fee revenue.

There is considerable debate on how an unbiased search engine in an inherently information-asymmetric world would look like[67]. The "conduit" theory sees search engines as passive intermediaries that make an "objective" selection of relevant search results in response to a user's search query. The ideal consumer-focused search engine would be a "trusted advisor" that presents results that match his preferences. That search engine would frustrate the preferences of service suppliers on the platform, and its own profit-maximizing objective. The appeal for "search neutrality" can be situated in that context. The "editor" theory sees search as a subjectively curated ranking of results in response to a query, with the search engine as an active editor. The editor view implies that there is no such thing as search neutrality because any ranking represents the search engine operator's profit maximizing view. In reality, search results are necessarily a combination of objective conduit and subjective editing. De los Santos et al[68] demonstrate how search operators are squeezed between the wishes of different types of platform users and carve out a profit margin while keeping all parties reasonably but not entirely satisfied. The stronger their market position, the more they may distort the information picture. Locked-in users have no choice to go elsewhere for their services. Competitive pressure may sometimes limit platforms' margin for manoeuver[69]. These models show how

62 Hagiu, Teh and Wright, 2020 Should Amazon be allowed to sell on its own marketplace? Feng Zhu and Qihong Liu, Competing with Complementors: An Empirical Look at Amazon.com.

63 Bergemann, Bonatti and Gan (2020), op.cit. In this model, data collection for advertising has a negative effect on the welfare of data originators because there is no compensatory service offered in return for the data.

64 See https://ec.europa.eu/commission/presscorner/detail/en/IP_19_4291

65 Hagiu, Teh and Wright, 2020 Should Amazon be allowed to sell on its own marketplace? Feng Zhu and Qihong Liu, Competing with Complementors: An Empirical Look at Amazon.com.

66 See for example Babur de los Santos and Sergei Koulayev (2017) Optimizing Click-through in Online Rankings with Endogenous Search Refinement, Marketing Science, 36(4): 542–564 2017.

67 Grimmelmann (2012) Some Skepticism About Search Neutrality, in The Next Digital Decade: Essays on the Future of the Internet; Grimmelman (2014) Speech Engines, 98 Minnesota Law Review 868 (2014)

68 Op. Cit.

69 Strucke and Ezrachi (2017) When competition fails to optimize quality: a look at search engines. Yale Journal of law and technology, vol 18(1).

ranking bias is inherent to the platform's use of asymmetric information. Platforms need to drive a wedge between the preferences of users on different sides of the market in order to extract a profit margin to ensure the sustainability of their business model. More recent information theory models expand this insight from rankings to the quality of information collected and shared by platforms[70].

Note that not-for-profit platforms would not perform better in this respect. They could limit their financial needs to cost recovery and charge a fixed fee to users, possibly in function of their intensity of use. The side of the market that pays the fee would receive the most optimal information to match their preferences. Other sides may still suffer from bias in the collection and use of information. A platform cannot use its data to simultaneously maximize the welfare of all users on all sides of the market, unless their preferences would be perfectly aligned. Information asymmetry is a fact of life in digital platform economies.

Independent service providers have alleged that the "Extended Vehicle" data governance model preferred by many vehicle manufacturers may lead to self-preferencing in vehicle data markets[71]. Under this model, vehicle manufacturers have the only direct access to all data collected by connected vehicles; aftermarket service providers can only get access to these data via a back-end server under the control of the manufacturer. Independent service providers claim that this may distort competition. Competition policy tools can be used to address individual cases where a lack of access to data forecloses such providers. The rise of digital car technology has shifted focus on the EU Motor Vehicle Type Approval Regulation (2018) as a regulatory tool to define an appropriate level of information fine-graining to restore information symmetry between authorised and independent service providers. Industry self-regulation has failed because of weak incentives for industry players to come to an agreement.

This example brings us to data sharing that is often touted as a means to overcome information asymmetry and maximize social welfare benefits for society[72] because it generates economies of scope in re-use. Data sharing markets may fail however when the data originator or collector perceives a risk of negative repercussions on his private welfare[73]. Data-driven platforms may offer compensation for this perceived risks, for instance by offering consumers a free service in return for sharing their data, or offering firms enhanced market access in return for sharing their data. Alternatively, platforms can modulate the degree of fine-graining and segmentation of the data they collect and share. Mandatory data sharing upsets these platform strategies, both on the data collection and the data use side of the platform. It may result in less data collection and undermine the positive externalities from data aggregation. Data policy makers need to carefully balance these positive and negative aspects of data-driven platforms.

3.6 Missing markets because of high transaction costs and risks

High transaction costs in the analogue economy prevented the emergence of many types of markets. Digital data massively reduce information costs and thereby facilitate market entry for consumers and small suppliers, from small hotels and bed & breakfasts that can now compete with large hotel chains on accommodation booking platforms, to independent taxi drivers who can offer their services on Uber and Lyft, and people entering the online labour market, or staying in touch with a large number of family, friends and professional contacts on social media. All this is made possible by intermediary online platforms. Markets that were "missing" in the pre-digital era suddenly emerge as a result of declining market entry and transaction costs. However, even in the digital data economy some markets still remain blocked because of high transaction costs. Moreover, new services are required in order to keep digital markets running but they may not appear because of still too high transaction costs and risks. In this section we present a few examples of missing markets because of high transaction costs and risks, and explore how these market failures may be addressed by a mixture of regulatory intervention and private third-party intermediation.

70 Bergemann, Bonatti and Gan (2020) op.cit.
71 See Martens and Muller-Langer, op.cit.
72 OECD (2016), op.cit. OECD (2019), Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies, Paris, November 2019.
73 See Bertin Martens & Alexandre de Streel & Inge Graef & Thomas Tombal & Nestor Duch-Brown, 2020. "Business-to-Business data sharing: An economic and legal analysis," JRC Working Papers on Digital Economy 2020-05, Joint Research Centre.

3.6.1 Transaction costs in personal data services markets

Personal data are an example of market failure because of high transaction costs and missing markets for services that could reduce these costs. Under the EU GDPR, data subjects have the right to consent to the use of their personal data before a firm can collect their data. This gives rise to frequent popping up of consent notices when consumers browse the internet. Consumers rarely read these consent notices however because the time involved is often not worth the effort. Even if they do they find it hard to make sense of the notices and understand what will happen to their personal data if they give their consent. Personal data consent notices are difficult to read and uninformative about possible data re-use[74]. A consumer survey confirms consumers' ambiguous attitudes towards privacy notices[75]. Another recent consumer survey[76] illustrates how risk assessments about sharing personal data on the internet vary widely according to type of data. Financial and biometric information command high subjective opportunity costs. Data use for advertising purpose are not perceived as entailing a significant privacy cost. Location and social network are somewhere in the middle. Valuations vary across countries and by gender and age.

More than two decades of research into the economics of privacy[77] have not produced an objective measure of the opportunity costs of personal data sharing. The re-use of personal data has ambiguous welfare effects. It can increase personal welfare when the data are re-used, for example, by search engines to reduce search costs and provide better search results that are more in line with consumer preferences. It may reduce welfare when data are re-used for targeted advertising that is more persuasive than informative and drives consumers away from their original preferences. Informative advertising helps consumers to make better choices. Persuasive advertising diverts consumer attention away from their preferences. This is a long-standing and unresolved debate in the marketing literature. If academics cannot solve the debate, consumers have an even harder time to assess the privacy costs of sharing personal data with an online service provider.

High transaction costs make the current system of consent notices dysfunctional, especially in the presence of depressed market values for private data. What seems to be missing is an active market for privacy management services. Many private start-ups have tried to enter the market for Personal Information Management Services (PIMS)[78]. They offer an intermediary platform to handle personal data exchanges with commercial platforms. However, none of these have scaled-up to become significant market players in personal data markets. The reason is clear: they do not really reduce high individual transaction costs[79]. Management costs are still relatively high, at least in time spent on the platform, compared to the depressed value of individual personal data. That makes their services unattractive to consumers who still overwhelmingly prefer dis-intermediated direct data exchange with platforms, clicking almost blindly on the consent notice.

More economically feasible personal data management requires technology that substantially lowers transaction costs. This could happen for example when consent notices become standardised and machine readable so that they can be processed by AI-driven machines. Standardisation would include the identity of the data collector, the purpose for which it is collected, the level of fine-graining in use of the data, and third-party commercial partners that may access the data. A privacy service provider could automatically link to these third-party privacy protocols and estimate possible risks for the data subject in function of his or her use of the internet. The service provider could then grant or deny consent in function of pre-set consumer preferences. Machine learning could gradually improve the efficiency by learning from consumer behaviour and across individuals and websites and collecting a detailed map of data sharing practices between firms and websites to suggest alternative service providers with lower privacy costs. Automation of the consent process would complete it in milliseconds, saving data subject a substantial amount of time. The bottleneck lies in the standardisation process however. Platforms can produce their own standardised consent notice but

[74] See for example Fred H. Cate, Viktor Mayer-Schonberger, Notice and consent in a world of Big Data, International Data Privacy Law, Volume 3, Issue 2, May 2013, Pages 67–73

[75] The survey confirms that nearly two thirds of consumers would appreciate government intervention in setting privacy rules but only about 20% of consumers bother to regularly read privacy notices. Results from the Brookings survey can be found here: https://www.brookings.edu/blog/techtank/2019/05/21/brookings-survey-finds-three-quarters-of-online-users-rarely-read-business-terms-of-service/

[76] Jeffrey Prince and Scott Wallsten (2020) How Much is Privacy Worth Around the World and Across Platforms? Technology Policy Institute, Washington DC, January 2020.

[77] Alessandro Acquisti, Curtis Taylor, and Liad Wagman (2016), The economics of privacy, Journal of Economic Literature 2016, 54(2), 442–492.

[78] Mydata.org (https://mydata.org/) is one example among many initiatives to help individuals manage their personal data. The European Data Protection Supervisor has advocated the use of Personal Information Management Systems (PIMS).

[79] Jan Kramer (2020) Personal data portability in the platform economy, Journal of Competition Law and Economics. Jan Kramer, Pierre Senellart and Alexandre de Streel (2020) Making data portability more effective for the digital economy, CERRE report, Brussels, June 2020.

without interoperability the system would run into high obstacles. Collective action seems to be required and that requires regulatory intervention[80].

3.6.2 Transaction costs and lack of transparency in advertising markets

A similar lack of transparency management services occurs in online advertising markets[81]. Online advertising can be split between "walled gardens" in search (Google) and social media (Facebook), and open display advertising where Google holds a strong position. Advertising is a two-sided market between publishers and advertisers, with several layers of intermediary platforms that do intermediate matching and price auctions for the supply of ad publishing windows and the stock of ads produced by advertisers. For every euro spent on ads by the advertiser, only 62 cents reaches the publisher, the rest remains in intermediate steps, largely dominated by Google[82]. It is challenging for advertisers to verify publishing and views of ads because of lack of transparency in intermediate stages. Price auctions in these market are problematic[83] because Google itself participates in the bidding while it has privileged information on the offers of its competitors. Self-preferencing is an issue. Data transparency and sharing through open standards and automated ad market tracking tools could be the solution. It could improve transparency and oversight for advertisers, publishers and content providers, increase competition and enable all participants to get a better view on what they pay for and what they achieve. It would create an ad data services market. Increased data sharing and transparency in the advertising eco-system may run into all kinds of adverse effects though, not least because it tests the limits of consumer privacy and commercial confidentiality.

3.6.3 Risks

There are circumstances where potential data suppliers refrain from participation in the production of services markets because it may be costly for them. An example is pooling mobility data between transport service providers in a city. This can have positive social welfare effects by improving traffic management and reducing congestion and pollution. Carballa[84] demonstrates how commercial transport service providers (buses, metros, taxis, e-scooter platforms, etc) may gain or lose market shares if they agree to share their data on a common platform, depending on price, substitutions and effects. Competitors may use the data to improve their offers and increase their market share. Alternatively, being on the common platform may attract more users to a particular provider. The net impact is an empirical question. If the net impact were negative, transport providers would have no incentive to participate in the platform. The platform may be in a position to compensate losers by re-allocating part of the overall social welfare surplus to them. For example, if drivers are willing to pay a positive price for improved congestion management, some of that revenue could be re-allocated to transport service providers that lose from participation. Withdrawing from the platform will reduce social welfare for all however. Regulators may have to intervene to make data sharing mandatory to overcome coordination problems, in the public social welfare interest[85].

Another dimension of transaction costs is ex-post risk in the execution of contracts. According to incomplete contract theory contracts are necessarily of finite length and can never include provisions for all possible events. Contracts inevitably come with residual uncertainties that can give rise to ex-post costs during monitoring and execution of a contract. This is especially the case for trade in non-rival and hard-to-exclude

---

[80] Posner and Weyl (2019) propose a particular variant on this theme. They suggest that data subjects should unite in unions to negotiate a higher value for their data with data collecting platforms. Automated data consent notices would reduce coordination and market entry costs for such unions. These unions would still face the problem of allocating the social value of the data between private members. See also the conclusions section in Bergemann, Bonatti and Gan (2020).

[81] The UK Competition and Markets Authority (CMA) conducted a detailed study on online advertising markets. This paragraph was inspired by that report. See https://assets.publishing.service.gov.uk/media/5efc57ed3a6f4023d242ed56/Final_report_1_July_2020_.pdf

[82] Damien Geradin and Dimitrios Katsifis Google's (Forgotten) Monopoly: Ad Technology Services on the Open Web, TILEC Discussion Paper, 21 May 2019. Spark Ninety, Transparency in programmatic online display advertising markets; the European Commission's Platform Observatory, Jan 2020.

[83] Geradin and Katsifis (2019); op.cit. Several EU competition authorities have launched investigations in online advertising, including the UK, FR and DE. A UK CMA study is exploring potential remedies for ads markets that could be part of an ex-ante regulatory regime.

[84] Bruno Carballa-Smiechowski (2018) Determinants of coopetition through data sharing in mobility-as-a-service.

[85] The European Commission's initiative to promote business-to-government data sharing "in the public interest" should be seen in this context. See European Commission (2020) Towards a European strategy on business-to-government data sharing for the public interest, Report of the B2G expert group.

data. Dosis and Sand-Zantman[86] distinguish between contractible and non-contractible data rights. Some contractual provisions may be unenforceable, non-monitorable or lack a commitment device. As a result, contracts are subject to the hold-up problem: parties will try to re-negotiate the contract when an unforeseen or non-committable event occurs. This includes risks from data leaks, unexpected data quality problems or processing errors. In traditional contracts, unexpected costs and benefits are assigned to the owner of the traded good or service. In the absence of legal ownership rights[87], data are ruled by de facto exclusive control. Data holders can use technical protection measures to ensure their exclusive use of the data. They may grant use rights to other parties through bilateral contracts. These contracts only bind the contracting parties, not third parties. In case of data leaks, the data holder has no leverage over third parties that might get hold of the data – except in cases where data benefits from intellectual property protection under the copyright, sui generis database or trade secret regimes, which offer protection against, respectively, reproduction, re-utilisation and unlawful acquisition or use of the protected subject matter. These risks may reduce data collectors' incentives to make data available for re-use and be more restrictive in granting data access. The risks of contractual hold-up may be too big for holders of valuable or commercially sensitive datasets. The Facebook – Cambridge Analytica case has amply demonstrated the risks of bilateral contracting for non-rival data.

Some authors have suggested assigning data ownership rights to overcome this problem[88]. From an economic perspective, ownership rights are residual rights: the costs and benefits of events that are not foreseen in a contract or law are automatically allocated to the owner of the residual rights. Debates on the possible introduction of such rights[89] have faltered and attention has now shifted to introducing data access rights[90]. Ownership and access rights are complements. Who should get such rights, if any, is not an easy question. For personal data, there is a "natural" rights holder, the data subject. For non-personal machine-generated data that may involve several parties for the co-generation of the data, it is often hard to unambiguously identify a "natural" rights holder. For example, in agriculture land owners, land operators, machine manufacturers, machine operators, sensor owners, data analytics providers, etc. may all claim rights over the data[91].

A more pragmatic solution may be for data exchanging parties that perceive high post-contractual risks in the execution of the exchange to appoint a neutral third-party intermediary who is tasked with managing the exchange in accordance with the terms of a contract. For example, a mobility service provider in a city may require pooled data from all mobile phone operators in that city to create detailed insights on citizen mobility patterns. None of the data suppliers trust the other to handle the data pool that has strategic commercial value for competitors. Solving this coordination problem requires a trusted third-party intermediary who collects the data, performs the analysis and ensures that only the processed results are shared with agreed users. This is the domain of semi-commons or governance agreements that seek to overcome the pitfalls of commons – that lead to overutilization and underinvestment and facilitate free-riding – and anti-commons – exclusive private use that leads to underutilization and keeps data locked in silos[92]. Semi-commons are often costly to manage. They are economically feasible when the value of the agreement for the participants exceeds the costs.

Data trusts and industrial data platforms fit the neutral profile[93]. In order to guarantee enforcement, the intermediary should be neutral and have no stake in the data or the outcomes of the analysis. That avoids strategic behaviour at the expense of the participants. The intermediary should only receive a fixed remuneration to produce the desired outcome. This permits him to act credibly as a trusted service provider for contractual commitments. He can enforce the commitment because he has full control over the data and access to the server. That reduces post-contractual risks and monitoring costs for the participants. Commercial for-profit data platforms may also provide guarantees against data leaks but they will exploit the

86 Anastasios Dosis and Wilfried Sand-Zantman (2019) The Ownership of Data, April 2019
87 Duch-Brown et al., 2017, op. Cit.
88 See Zegg (2016) and Weibe (2017)
89 European Commission (2017) Communication on "Building a European Data Economy" (COM(2017) 9 final) and Staff working document on the free flow of data and emerging issues of the European data economy (SWD (2017) 2/1.
90 European Commission (2018) "Towards a common European Dataspace", COM(2018) 232 final. Jozef Drexl (2018) Data access and control in the era of connected devices, BEUC Brussels, January 2018.
91 Can Atik and Bertin Martens, Competition Problems and Governance of Non-personal Agricultural Machine Data: Comparing Voluntary Initiatives in the US and EU, Digital Economy working paper 2020-07, Joint Research Centre of the European Commission, Seville.
92 Henry Smith (2005) Governing the tele-semi-commons, Yale Journal of Regulation, vol 22(289). Exclusive private property rights are cheaper to manage – the exclusive owner sets the price – and so are full commons because the price falls to zero.
93 Alexandre De Streel and Thomas Tombal, 2019, pp 11 The Fifty Shades of Data Sharing and the Law; University of Namur, CRIDS/NADI, CERRE.

data in their own interest, and sometimes against the interests of the data providers. They create new sources of ex-post risks.

4 Data-driven ecosystems

So far we discussed the use of data within a single platform. In this section we consider the use of data across platforms and markets. Information from one market or platform can be re-used in another market.

In classic economics, firms that re-use resources obtained in one market to enter into another market are known as conglomerates[94]. Conglomeration can be fuelled by excess resource capacity and market power considerations. Re-use of financial, human and physical resources within a firm to enter a different market may be interpreted variously as leading to greater efficiency or as a sign of inefficient markets for shareable inputs[95]. When capital markets are not functioning well, firms with excessive financial resources may decide to invest directly in other sectors rather than investing their surplus through capital markets. Firms may also cross-subsidize between different units to expand market power. Google advertising revenue subsidizes many of its other activities. Transmission mechanism may include data. Contrary to physical goods, data are intangible resources that are non-rival by nature and can be re-used for different purposes at the same time, without diminishing their value for any of these purposes. Moreover, a unique feature of data is that economies of scope in data aggregation can generate additional value from combining data across different markets and platforms.

Apart from creating a single legal and financial conglomerate structure, firms can also decide to collaborate through alliances that exploit complementarities between products and services. Such alliances fall outside collusion prohibitions as long as markets are seemingly unrelated. An ecosystem can be defined as a set of firms that coordinate complementarities in different markets in a deliberate non-generic and strategy way with a view to create more value[96]. Examples of digital ecosystems include operating systems, including apps operating systems. While it is not hierarchical or subject to vertical integration, there is an element of market power that makes it non-voluntary. It creates mutual dependencies. Their distinctive feature is that they provide a structure within which complementarities (of all types) in production and/or consumption can be contained and coordinated without the need for vertical integration. Powerful firms craft rules and shape the process of ecosystem development to tie in complements and make complementors abide by them.

Data can be leveraged as a transmission channel to coordinate complementarities. For example, Google combines data from search, e-mail, apps, location, maps, etc. to drive its advertising business. A key feature of this ecosystem is standardisation of data interfaces between these products so that data can easily be transferred and interpreted across the different components. For example, smartphone hardware manufacturers who wish to use Android apps are required to join the Open Handset Alliance which obligates members to use only Google approved Android versions. In this way, even though Android is open source, Google's control prevents fragmentation of the code base by means of some level of standardization. The downside is that potential operating system innovations are not interoperable with Google data services and Google may be able to charge higher prices for those services. The offsetting benefit is that app developers and hardware manufacturers have to contend with fewer variants of the Android operating system than they otherwise would and are thus able to ensure interoperability.

Very large platform firms, such as the GAFAMs (Google, Apple, Facebook, Amazon and Microsoft), have become sprawling businesses that keep expanding into new sectors and domains, re-using data and algorithmic tools and applying "envelopment"[97] and bundling strategies to enter into new markets. Through envelopment, a provider in one platform market can enter another platform market, combining its own functionality with the target's in a multi-platform bundle that leverages shared user relationships. The envelopment hypothesis builds on the traditional view of bundling and extends this to include the strategic management of a firm's user network. Envelopers capture share by foreclosing an incumbent's access to users. In doing so, they harness the network effects that previously had protected the incumbent. Eisenmann et al (2011) distinguish three envelopment strategies: (a) supply side economies of scope: components and services that can be re-used in other platforms, (b) demand side economies of scope: user overlaps that can be leveraged and (c) negative price correlations between different services. Data can be an important vector for the first two strategies.

---

94 Boureau Marc and Alexandre De Streel (2019) Digital conglomerates and EU competition policy, Paris Telecom and Université de Namur, March 2019.

95 David J.. Teece, 'Economies of Scope and the Scope of the Enterprise' (1980) 1 Journal of Economic Behavior and Organization 223; David. Teece, 'Towards an Economic Theory of the Multip-Product Firm' (1982) 3 Journal of Economic Behavior and Organization 39.-63; John. C. Panzar and Robert. D. Willig, 'Economies of Scope' (1981) 71 The (2) American Economic Review 268.

96 M. Jacobides, C. Cenammo, A. Gawer "Towards a theory of ecosystems". Strategic Management Journal, May 2018.

97 S. Eisenmann, G. Parker and M. Van Alstyne (2011) "Platform envelopment strategies", Strategic Management Journal, vol 32(12).

Another envelopment strategy uses privacy policy statements in different platforms to combine personal data[98]. Data can be collected in one market and re-used in another market. A good example are the Facebook/Whatsapp and the Facebook/Instagram mergers. Data re-use from apps and browsers in online advertising are also examples. Platform owners sometimes enter complementors' product spaces and compete against them. Research found that Amazon uses data collected on its e-commerce platform to decide on market entry[99]. It is more likely to target successful product spaces and less likely to enter product spaces that require greater seller efforts to grow, suggesting that complementors' platform-specific investments influence platform owners' entry decisions. While Amazon's entry discourages affected third-party sellers from subsequently pursuing growth on the platform, it increases product demand and reduces shipping costs for consumers.

98 Condorelli, Daniele and Padilla, Jorge, Harnessing Platform Envelopment in the Digital World (December 14, 2019).
99 Zhu and Liu (2018) Competing with Complementors: An Empirical Look at Amazon.com.

5 Concluding remarks

The data economics issues that we discussed here have much in common with the law and economics of intellectual property rights (IPR), such as patents and copyrights. The economic characteristics of data, non-rivalry and no natural excludability, are similar to those of innovation. IPR give exclusive ownership rights to innovators in order to yield a return on investment and an incentive for innovation. IPR policies struggle with the same balancing act as data policies, between the social welfare costs of monopolistic exclusive rights and the social welfare gains from the innovation incentive effects. Monopolistic IPR license pricing, above the marginal cost of reproduction, reduces access to innovation. This is an unavoidable social harm in order to generate dynamic innovation benefits. Society manipulates that balance by limiting the scope of exclusive rights. Similar considerations apply to data collection, access and use, including in online platforms. It took several centuries for society to develop a coherent system of IPR rights and it is still evolving, driven by technology that affects the cost of innovation production and dissemination and therefore the balance between protection and access. Digital data are a very new product in society. There are lively discussions between proponents of exclusive ownership rights and defendants of more open access rights[100]. A major difficulty with data is the attribution of such rights. Innovations are usually produced by a well-defined innovator or group of innovators with common interests. Data by contrast require at least two parties, a data originator and a collector, and often more, sometimes with diverging interests. While personal data rights may be "naturally" attributed to a data subject, attribution is more difficult for non-personal data where many parties may be involved in origination, collection, aggregation and analysis of the data. Changes in attribution of rights may affect entire data value chains and downstream services markets. They will affect the pace of innovation that data can bring to society.

More importantly, both ownership and access rights overlook the inherent social value of data and the externalities that they entail[101]. A single data originator or collector is usually not in a position to internalize these externalities. Market failures will remain. The debates often give the impression that the attribution of exclusive rights and access rights or data sharing rights are policy objectives in themselves. This paper has emphasized that such rights are only policy instruments that should be used to maximize the social welfare that society as a whole can derive from the use of data.

In this paper we focused mainly on social welfare as a benchmark for identifying market failures and policy intervention. Public policy economics defines social welfare measure[102] as the combined welfare of all stakeholder groups in society, consumers and producers. However, the mainstream benchmark in competition law is a narrower consumer welfare benchmark[103]. These two measures can easily lead to contradictory conclusions. For example, regulatory intervention to open market access on one side of a platform may reduce welfare on other sides of the platform market. Classic economics rejects the comparison of welfare gains and losses between groups or individuals because consumer welfare is assumed not to be quantifiable. Alternative approaches accept quantification but open the door to measures of social welfare improvement whereby some parties gain at the expense of others. Economics distinguishes between strictly Pareto-improving welfare measures whereby no agent loses welfare, and a less stringent Kaldor-Hicks[104] welfare measure whereby some agents may lose but could be compensated by the gains that other agents make in order to avoid equity concerns. Western societies have historically put emphasis on individual wellbeing and are reluctant to impose private costs on individuals in order to achieve wider social welfare gains, unless they are compensated by transfers to ensure some degree of equity. Other societies have a more collective view of social welfare and attach less importance to individual welfare. They would find it easier to accept private costs as long as overall welfare increases. This underscores the borderline between the economics of data and cultural, social and political value judgements in society on how to maximize societal welfare from data.

100 The European Commission's Communications (2017, 2018, 2020, op.cit.) on data issues over the last years reflect this societal debate. See also 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; 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.

101 Bergemann, Bonatti and Gan (2020), op.cit.

102 The notion of social welfare can be stretched beyond pure economics and include broader societal issues such as equity and income distribution, the protection of vulnerable groups, minorities, culture and political systems. See European Commission, Better Regulation Guidelines (2017).

103 Furman et al, 2019, op. cit.

104 For more details on Kaldor-Hicks measures, see for example https://en.wikipedia.org/wiki/Kaldor–Hicks_efficiency

References

Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian and Asuman Ozdaglar, Too much data: prices and inefficiencies in data markets; NBER Working Paper 26296, September 2019.

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman (2016), The economics of privacy, Journal of Economic Literature 2016, 54(2), 442–492.

Agrawal, Ajay, Avi Goldfarb, and Joshua Gans (2018), Prediction Machines: The Simple Economics of Artificial Intelligence, Harvard Business Review Press, 2018.

Aridor, Guy and Che, Yeon-Koo and Salz, Tobias, The Economic Consequences of Data Privacy Regulation: Empirical Evidence from GDPR (January 29, 2020). Available at SSRN: https://ssrn.com/abstract=3522845

Atik, Can and Bertin Martens, Competition Problems and Governance of Non-personal Agricultural Machine Data: Comparing Voluntary Initiatives in the US and EU, Digital Economy working paper 2020-07, Joint Research Centre of the European Commission,

Bajari, P and V Chernozhukov, A Hortaçu and J Suzuki (2018) The impact of big data on firm performance, an empirical investigation, NBER working paper nr 24334.

Bakos, Yannis and Erik Brynjolfsson (2000) Bundling and Competition on the Internet, Marketing Science vol. 19, No. 1, Winter 2000, pp. 63–8.

Bergemann, Dirk and Alessandro Bonatti, 2018, Markets for information: an introduction, CEPR discussion paper DP1314, 2018.

Bergemann, Bonatti and Gan (2020) The economics of social data. Yale university, Cowles Foundation discussion paper nr 2203 revised.

Boureau Marc and Alexandre De Streel (2019) Digital conglomerates and EU competition policy, Paris Telecom and Université de Namur, March 2019.

Bresnahan Tim and M Trajtenberg (1995) General purpose technologies: engines of growth, Journal of Econometrics Volume 65, Issue 1, January 1995, Pages 83-108

Brynjolfsson, Eric, Avinash Collis, W. Erwin Diewert, Felix Eggers, Kevin J. Fox (2019) GDP-B: accounting for the value of new and free goods in the digital economy. NBER Working Paper 25695, March 2019.

Caillaud, B., & Jullien, B. (2003). Chicken & egg: Competition among intermediation service providers, The RAND Journal of Economics, Vol. 34, No. 2 (Summer, 2003), pp. 309-328

Carballa Smiechowski, Bruno (2018) Determinants of coopetition through data sharing in mobility-as-a-service.

Cate, Fred and Viktor Mayer-Schonberger, Notice and consent in a world of Big Data, International Data Privacy Law, Volume 3, Issue 2, May 2013, Pages 67–73

Chiou L and C Tucker (2017) Search engines and data retention, implications for privacy and antitrust. NBER working paper nr 23815.

Claussen, Jörg and Peukert, Christian and Sen, Ananya, The Editor vs. the Algorithm: Targeting, Data and Externalities in Online News (June 5, 2019).

Coase, Ronald (1937) The nature of the firm, Economica, November 1937.

Crémer, J., Y-A de Montjoye and H. Schweitzer (2019) Competition policy for the digital era, Report for Commissioner Vestager, European Commission.

De Cornière and Taylor (2020) "Data and competition: a general framework with applications to mergers, market structure and privacy policy". Mimeo, February 2020.

European Commission (2020) Report of the Expert Group on B2G data sharing, available at https://ec.europa.eu/digital-single-market/en/news/meetings-expert-group-business-government-data-sharing

De los Santos, Babur and Sergei Koulayev (2017) Optimizing Click-through in Online Rankings with Endogenous Search Refinement, Marketing Science, 36(4): 542–564 2017.

De Streel, Alexandre and Thomas Tombal, 2019, pp 11 The Fifty Shades of Data Sharing and the Law; University of Namur, CRIDS/NADI, CERRE.

Dengler, Sebastian and Prufer, Jens, Consumers' Privacy Choices in the Era of Big Data (April 9, 2018). TILEC Discussion Paper No. 2018-014. Available at SSRN: https://ssrn.com/abstract=3159028

Dosis, Anastasios and Wilfried Sand-Zantman (2019) The Ownership of Data, April 2019

European Commission (2017) Communication on "Building a European Data Economy" (COM(2017) 9 final) and Staff working document on the free flow of data and emerging issues of the European data economy (SWD (2017) 2/1.

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

Drexl, Jozef (2018) Data access and control in the era of connected devices, BEUC Brussels, January 2018.

Eisenmann, Thomas, Geoffrey Parker, Marshall Van Alstyne (2011) Platform envelopment, Strategic Management Journal, Volume 32, Issue12, December 2011, Pages 1270-1285.

European Commission's "Better Regulation Guidelines" (2017) available at https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how/better-regulation-guidelines-and-toolbox_en

European Commission (2018) "Towards a common European Dataspace", COM(2018) 232 final.

European Commission (2020) "Data Strategy", available at https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

European Commission (2020) Towards a European strategy on business-to-government data sharing for the public interest, Report of the B2G expert group.

European Union, Directive 96/9/EC of the European Parliament and the Council of the European Union, 11 March 1996, on the legal protection of databases.

European Union, Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market

European Union, Regulation (EU) 2016/679 of the European Parliament and of the Council of 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.

European Union, Second Payment Services Directive (EU) 2015/2366.

European Union Regulation (EU) 2019/1150 on promoting fairness and transparency for business users of online intermediation services.

Franck Jens-Uwe and Martin Peitz (2019), Market definition and market power in the platform economy, CERRE, Brussels, May 2019.

Furman, J., D. Coyle, A. Fletcher, D. McAuley and P. Marsden (2019), Unlocking digital competition, Report of the digital competition expert panel, March 2019.

Gans, Joshua, The Specialness of Zero, Jan 2020, available at SSRN: https://ssrn.com/abstract=3486964

Geradin, Damien and Dimitrios Katsifis Google's (Forgotten) Monopoly: Ad Technology Services on the Open Web, TILEC Discussion Paper, 21 May 2019.

Grimmelmann (2012) Some Skepticism About Search Neutrality, in The Next Digital Decade: Essays on the Future of the Internet;

Grimmelman (2014) Speech Engines, 98 Minnesota Law Review 868 (2014)

Hagiu, A., & Wright, J. (2015). Marketplace or Reseller? Management Science, 61(1).

Hagiu, Andrei and Julian Wright (2020), When data creates competitive advantage, Harvard business review, January 2020.

Iansati, Marco and Karim Lakhani (2020) Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World, Harvard Business Press.

Jacobides,Michael, Carmelo Cenammo, Annabelle Gawer (2018) Towards a theory of ecosystems, Strategic management journal, March 2018.

Jones, Charles and Christopher Tonetti, Nonrivalry and the economics of data, NBER Working Paper nr 26260, September 2019.

Kerber, Wolfgang (2019) Data Governance in Connected Cars: The Problem of Access to In-Vehicle Data, Journal of intellectual property and information technology, vol 9(3) 2018.

OECD Maximizing the economic and social value of data, understanding the Benefits and Challenges of Enhanced Data Access, Directorate for Science and Technology, Committee on Digital Economic Policy, Paris, November 2016.

OECD (2019), Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies, Paris, November 2019.

Martens, Bertin (2016) "An Economic Policy Perspective on Online Platforms," JRC Working Papers on Digital Economy 2016-05, Joint Research Centre.

Martens B. and F. Müller-Langer (2020) Access to digital car data and competition in aftersales maintenance markets, Journal of Competition Law and Economics, 2020.

McAfee, Preston, Justin Rao, Aadharsh Kannan, Di He, Tao Qin, Tie-Yan Liu (2015), Measuring Scale Economies in Search, Microsoft slides.

McNamee, Roger (2019) Zucked: waking up to the Facebook catastrophe.

Neumann, N. C Tucker and T Whitfield (2018) How effective is black box digital consumer profiling and audience delivery? Evidence from field studies. Mimeo, MIT.

Palfrey, John and Urs Gasser, Interop: The Promise and Perils of Highly Interconnected Systems, 2012.

Panzar, John and Robert D. Willig, Economies of Scope, the American Economic Review, Vol. 71, No. 2, May 1981.

Parker, G & Alstyne, M.. (2005). Two-Sided Network Effects: A Theory of Information Product Design. Management Science, 51(10), 1494-1504.

Parker, Geoffrey and Van Alstyne, Marshall W. and Jiang, Xiaoyue, Platform Ecosystems: How Developers Invert the Firm (August 17, 2016). Boston University Questrom School of Business Research Paper.

Parker, Geoffrey and Van Alstyne, Marshall W. and Jiang, Xiaoyue, Platform Ecosystems: How Developers Invert the Firm (August 17, 2016). Boston University Questrom School of Business Research Paper.

Posner and Weyl, 2019 Radical markets

Prince Jeffrey and Scott Wallsten (2020) How Much is Privacy Worth Around the World and Across Platforms? Technology Policy Institute, Washington DC, January 2020.

Prüfer, Jens and Christoph Schottmüller (2017), Competing with Big Data, mimeo, Tilburg University.

Reimers, Imke and Joel Waldfogel (2020) Digitization and Pre-Purchase Information: The Causal and Welfare Impacts of Reviews and Crowd Ratings, NBER working paper nr 26776, February 2020.

Ribers, Michael and Hannes Ullrich, 2019. "Battling Antibiotic Resistance: Can Machine Learning Improve Prescribing?," Discussion Papers of DIW Berlin 1803, DIW Berlin, German Institute for Economic Research.

Rochet, J.-C., & Tirole, J. (2003). Platform Competition in Two-sided Markets. Journal of the European Economic Association, 1(4), 990-1029.

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

Rosen, Sherwin (1983) Specialisation and human capital, Journal of Labor Economics, Volume 1, Number 1 Jan., 1983.

Scott-Morton, F., P. Bouvier, A. Ezrachi, B. Julien, R. Katz, G. Kimmelman, A.D. Melamed, and J. Morgenstern (2019), Report of the Committee for the study of digital platforms, market structure and antitrust subcommittee, George Stigler Center for the study of the economy and the state, University of Chicago Booth School of Business, May 2019.

Smith, Henri (2005) Governing the tele-semi-commons, Yale Journal of Regulation, vol 22(289).

Strucke and Ezrachi (2017) When competition fails to optimize quality: a look at search engines. Yale Journal of law and technology, vol 18(1).

Spark Ninety (2020) Transparency in programmatic online display advertising markets; slide presentation to the European Commission's Platform Observatory, Jan 2020.

Teece, David (1980), Economies of scope and the scope of the enterprise, Journal of economic behaviour and organisation, 1980.

Teece, David (1982) Towards an economic theory of the multi-product firm, Journal of economic behaviour and organisation, 1982, pp 39-63.

Tucker, Catherine (2019), Digital data, platforms and the usual antitrust suspects: network effects ,switching costs, essential facility, mimeo, January 2019.

Ursu R, (2015) The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions.

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;

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.

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.

The European Commission's science and knowledge service

Joint Research Centre

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.