European Commission

JRC TECHNICAL REPORT

JRC Digital Economy Working Paper 2020-05

Business-to-Business data sharing: An economic and legal analysis

Bertin Martens
Alexandre de Streel
Inge Graef
Thomas Tombal
Néstor Duch-Brown

2020

Joint Research Centre

This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Contact information
Name: Bertin Martens
Address: Inca Garcilaso 3, 41092 Seville (Spain)
Email: Bertin.Martens@ec.europa.eu

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

JRC121336

Seville: European Commission, 2020

© European Union, 2020

The reuse policy of the European Commission is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (https://creativecommons.org/licenses/by/4.0/). This means that reuse is allowed provided appropriate credit is given and any changes are indicated. For any use or reproduction of photos or other material that is not owned by the EU, permission must be sought directly from the copyright holders.

All content © European Union, 2020.

How to cite this report: Bertin Martens, Alexandre de Streel, Inge Graef, Thomas Tombal, Néstor Duch-Brown, Business to business data sharing: an economic and legal analysis, Digital Economy Working Paper 2020-05, European Commission, Seville, 2020, JRC121336.

Contents

Executive summary 4

Introduction 10

1. The economic characteristics of data and data markets 12

1.1. The economic characteristics of data 12

1.2. The interaction between data collection, use and services markets 16

2. Private markets for B2B data sharing and market failures 19

2.1. Monopolistic data markets 19

(a) The substitution effect in secondary services markets 20

(b) Vertical integration 21

(c) The need for complementary inputs 22

(d) Economies of scope in data aggregation 23

2.2. Other causes of data market failures 25

(a) Externalities 25

(b) Incomplete contracts, risks, transaction costs and missing data markets 25

(c) Asymmetric and imperfect information 27

3. Third-party intermediaries to reduce market failures 28

3.1. Reducing risk 29

3.2. Transaction costs and standardisation 29

3.3. Internalisation and redistribution of externalities 32

3.4. The role of third-party intermediaries 33

4. Competition Law and Regulation to remedy market failures 35

4.1. *Ex post* Competition enforcement 35

4.2. *Ex ante* regulation 38

(a) Data portability rights for the data subject 40

(b) Mandatory data access right for other parties 45

5. Conclusions 49

References 51

Executive summary

The European Commission announced in its Data Strategy (2020) its intentions to propose an enabling legislative framework for the governance of common European data spaces, to review and operationalise data portability, to prioritise standardisation activities and foster data interoperability and to clarify usage rights for co-generated IoT data. This Strategy starts from the premise that there is not enough data sharing and that much data remain locked up and are not available for innovative re-use. The Commission will also consider the adoption of a New Competition Tool as well as the adoption of ex ante regulation for large online gatekeeping platforms as part of the announced Digital Services Act Package[1]. In this context, the goal of this report is to examine the obstacles to Business-to-Business (B2B) data sharing: what keeps businesses from sharing or trading more of their data with other businesses and what can be done about it? For this purpose, this report uses the well-known tools of legal and economic thinking about market failures.

The economic characteristics of data

A key economic characteristic of data is non-rivalry: many parties can use the same dataset for a variety of purposes without functional loss to the original data collector. Non-rivalry is the fundamental driver of economic welfare gains in data sharing or re-use: if one firm collects data that can be used for many purposes, society would benefit if other firms could access and use these data. An economic interpretation of non-rivalry revolves around **economies of scope** that occur when the same product is re-used for another purpose. A second source of economic efficiency gains in the use of data comes from economies of scope in data aggregation. When two datasets are complementary, more insights and economic value can be extracted from merging them, compared to keeping them in separate data silos. Economies of scope should be distinguished from economies of scale in data. Both contribute to improving the prediction accuracy of datasets.

Privacy and commercial confidentiality are important for the autonomy of private decision-making by firms and individuals and for extracting private value from these decisions. Hence, this points at the importance of exclusive data access and control. **Data are not excludable by nature**. They require technical and/or legal protection to create exclusive access for one party. From an economic perspective, data sellers should have exclusive control over their data. Uncontrolled access will drive prices down to the marginal cost of reproduction, often close to zero. Since the law provides few means for firms to assert exclusive control rights over data, firms often resort to *de facto* exclusivity by means of contractual or technical protection measures.

Data market failures

Data may have social value that exceeds their private value. Data collected from one group of persons may have predictive value for the behaviour of another group.

[1] This paper does not prejudge the on-going debate on these policy tools. It is based on currently available information in the published Inception Impact Assessments for these policy instruments. No decisions have been taken by the European Commission on how they will ultimately look like.

Aggregated datasets may have more value than segmented sets. This implies an **inherent market failure in exclusive private control over data** and the need to find data governance arrangements that find a balance between sharing and exclusivity.

While economies of scope in data re-use constitute an argument in favour of wider access and diffusion of data, economies of scope in aggregation underpin the economic benefits of data concentration in large pools. It constitutes an incentive for digital data firms to expand their activities to other data-related services markets and build conglomerates. It may strengthen their market power because it increases entry barriers for new firms and diminishes incentives for innovation. The impact of economies of scope in data aggregation on society is ambiguous. **Policy makers need to trade off the social welfare from increased productivity and innovation against the anti-competitive market concentration forces at work**.

Apart from the benefits there is also a cost to data sharing and re-use. Private firms may incur costs when they share data with parties that can harm their interests. They take data sharing decisions in function of the expected benefits and costs. The question for policy makers is **whether these private data sharing incentives and decisions maximise the welfare of society** as a whole. A market failure occurs when the private value of data remains below its social value. Bridging that gap may require policy intervention. Reducing the market power of data holders and facilitating more data sharing can be part of the solutions.

Thus, **data sharing is not an objective in its own right. It is a means to achieve higher social welfare**. Exclusively private data use may lead to underutilisation of data. Other parties could make good use of the data but have no access. Fully shared common data pools are subject to overutilization. When data are freely available to all, there is no incentive anymore to invest in data. Intermediate semi-commons data governance solutions are more likely to be optimal. Governance regimes, whether private or public, can be costly to manage however. Whether they are worth the cost depends on the added-value generated by the governance agreement.

While exclusive access and **monopolistic market positions** are necessary to extract value from data, they are frequently a source of market failure. Monopolistic data markets occur when a firm collects data for which there are no close substitutes. **Whether the firm will allow re-use of the data, inside the firm or by another firm, depends on the relationship between the markets for primary and secondary use**. (i) If the secondary use competes with the first use, for instance for the production of a very similar service, it may refuse re-use. (ii) If the secondary service is a complement or neutral with respect to the first, the firm has an incentive however to promote re-use, either in-house or by another firm. In the latter case it would result in data sharing. However, in practice, it is not always a priori clear whether data can be used to produce a complementary or a substitute service. This uncertainty may create obstacles for data sharing. Still, the monopolistic data firm can charge a monopoly price to access the data and apply price discrimination between re-users to maximise his profits. The firm may foreclose the market and have an exclusive deal with one firm to re-use the data, or buy up the other firm (vertical integration) to produce the second service in-house.

Besides monopolistic market positions, there are other causes of market failure. **Coordination problems may be an obstacle to re-use**. Re-use of the data for the production of another service may require complementary inputs that the data-holding firm does not have. It requires cooperation with another firm to procure such

complementary input. (i) If the latter is available in a competitive market there is no problem. (ii) If both the data and the complementary inputs are sold by monopolistic firms, an anti-commons coordination problem occurs that requires negotiations between the two parties. This may reduce data sharing. Unequal bargaining power between the parties may facilitate data sharing but generate equity concerns. Third party intermediaries may help to overcome coordination obstacles. If the complementary input is another monopolistic dataset, economies of scope in data aggregation may occur if the two monopolistic firms agree on data sharing conditions.

High transaction costs in concluding data sharing agreement may be another obstacle. Indeed, data sharing contracts are necessarily incomplete and can lead to privacy and security risks and unforeseen outcomes between the contracting parties or with third parties. In the absence of ownership rights on data, bilateral data contracts cannot be enforced against third parties. Data may spill into the public domain or end up in the hands of parties that can cause harm to the original data holder. Such externalities result in misalignment of incentives between data collectors and re-users.

Imperfect and asymmetric information between individuals or small firms and large data collectors, which are almost natural states in a data-abundant digital world, **distort efficient decision making**. Re-allocation of data access rights may affect market outcomes. Externalities may reduce the value of data protection. Data collected on the behaviour of one set of agents has predictive value for the behaviour of others. This reduces the marginal value of a single person's or firm's dataset and diminishes incentives for privacy protection.

Market-based solutions for data sharing

Solving market failures does not always require mandatory behaviours. Third-parties may act as intermediaries and apply new technologies and ways of organising markets in order to overcome market failures and enable transactions that were previously not feasible. They can be private, public or community organisations that are neutral with respect to data uses or they can be active stakeholder in the added-value generated by the data. The European Commission's Data Strategy (2020) calls them "**common data spaces**". From an economic perspective, they can reduce *ex-ante* transaction costs and *ex-post* contractual risks, overcome coordination problems, act as commitment devices and facilitate self-regulation between private agents, and set interoperability standards.

Risk reduction is required when none of the participants in a data pool wants other parties to have access to their primary data but still wants to benefit from the value of the pooled data. **Data trusts and industrial data platforms** fit this profile. The intermediary should be **neutral** and have no stake in the data or the outcomes of the analysis. Overcoming pre-contractual transaction costs requires a more active intermediary who has a stake in reaching B2B data deals between providers and re-users but has no stake in the contents of the data transfer. Third-party operated B2B data sharing platforms may be more successful in closed groups with known participants than in open-ended groups of users. Re-use may have negative externality effects on the data supplier, or they may miss opportunities.

Intermediary platforms may also **reduce transaction costs** because they can contribute to standardisation and interoperability of the datasets that are exchanged. Voluntary standard setting works when the participants have an incentive to contribute to a standard. Standards may fail to emerge when stakeholders have conflicting interests.

The absence of standards may also create entrepreneurial opportunities for firms that offer private solutions to overcome data market fragmentation and interoperability problems. Strong intermediary market players may impose market-driven standards.

Data holders and re-users may not be able to reach an agreement when they fail to capture (and monetise) the gains from data sharing or do not agree on the distribution of the gains. Third-party intermediaries may help to overcome these hurdles by offering a **business model that can capture and monetise the externalities, and by setting up a redistribution** mechanism. These intermediaries usually become **non-neutral** and more active third parties that have a stake in the value of the pooled data. They may put data input providers in a weak bargaining position because they cannot realise the economies of scope from data aggregation on their own. In some cases, the benefits of data pooling constitute non-excludable public goods, for example pooling of health data to improve public health services. This positive externality is difficult to monetise for an individual health service provider who therefore has no financial incentive to participate in a health data pool. Government intervention may be required to overcome this market failure.

Compulsory data portability and access to increase data sharing

When market-based solution, possible nudged by the State, do not solve market failures, a **more active intervention from the State imposing data portability and access** may needed provided the risks of regulatory failures are limited. Such intervention may take place ex-post with competition law or ex-ante with regulation.

Ex-post competition enforcement: If data market failures occur only occasionally, competition authorities may be well-placed to address the problem ex-post and on a case-by-case basis. There are two possible scenarios: mandating access to data where a dominant firm refuses to do so and correcting discriminatory prices or unfair conditions for access to data held by a dominant firm.

First, the so-called **essential facilities doctrine** is relevant in the case of refusal of access to data as a form of abuse of dominance under Article 102 TFEU. There are some restrictions however. A first condition is that the data should be indispensable for the production of the alternative service. However, in some cases non-rivalry and wide availability of close substitute data may render a dataset non-indispensable. Second, access to data for the purpose of producing a service in a market where the dominant firm is not yet active itself, falls outside the scope of the essential facilities doctrine as currently interpreted in decisions and case law. Yet, these are precisely the scenarios that are to be expected in data markets. It may be necessary to **adapt these conditions to the characteristics of data** to enable effective enforcement of data access under competition law.

Second, market failures may also originate when a dominant firm gives access to data. **Self-preferential access to data** may distort downstream services markets. Data pricing conditions may be unfair and require redress, for instance through the intervention of a neutral third-party intermediary who decides on a fair price. Unfair price discrimination may increase social welfare but raise equity concerns in society.

Once a case has clarified the circumstances under which a refusal to share data amounts to abuse under Article 102 TFEU, this may set a precedent to which market players in other sectors adapt their behaviour as well. However, **competition interventions take**

time and the harm may be done before the case can be resolved. This is especially true in "winner takes all" markets, where it becomes difficult to contest the incumbent's position once the market has tipped in its favour. Moreover, the **design and the monitoring of data access remedies are very complex** and require an effective governance framework. Thus, legislators may decide to set up ex ante regulation for data sharing.

Hybrid and ex-ante regulatory intervention: When market failures occur on a wider scale and on a regular basis, it might be necessary for legislators to set ex-ante mandatory rules that reduce the exclusive control of data holders. The envisaged **New Competition Tool would enable a hybrid form of intervention** (a mix between ex post competition enforcement and ex ante regulation), allowing competition authorities to impose remedies, including data access remedies, to address structural competition problems in a market without the need to establish a violation of the competition rules. The envisaged regulatory instrument for large online gatekeeper platforms would enable ex ante regulatory intervention in data markets. Legislators can impose obligations on data holders and **assign legally binding access rights to stakeholders in the data market**, ranging from full and exclusive ownership rights to data, to more specific and limited rights, such as access and portability rights. The choice of the parties to which these specific rights are allocated affects economic outcomes, both in upstream data markets and in downstream services markets.

The **EU already has several legislations in place**. The EU GDPR allocates specific rights to personal data subjects. There are no legally defined rights for non-personal data. The Database Directive, the Platform-to-Business Regulation, the Free-Flow of Data Regulation, may play a role in B2B data sharing. Consumer law, in particular the Digital Content Directive also plays an increasingly important role in allocating data rights to consumers. Besides the horizontal regulatory tools, there are sector-specific tools such as the Second Payment Services Directive, the Motor Vehicle Regulation and the Electricity Directive that affect data sharing in these sectors.

There are **two basic scenarios**. (i) When data originators have strong incentives to share data with alternative service providers, legislators may grant portability rights to the data originator to access and transfer data collected by a data holder. (ii) When data originators have no strong incentives, regulators can grant specific data access rights directly to data seekers.

With regard the first scenario, **portability rights exist only for personal data** under Article 20 of the GDPR. The data subject has an inalienable right to ask the data processor to retrieve his personal data or to transfer them to a third party. This creates a roundabout B2C2B channel to overcome obstacles to direct B2B data sharing in the case of personal data. It can be seen as a tool to facilitate switching and competition, and as a fundamental right to empower the data subject at the same time. However, **there are limitations**. It applies only to personal data actively provided or observed by the data subject. Data inferred or derived by the data controller are not covered by portability rights though they can constitute valuable personal data. Moreover, portability is not (yet) sufficiently operational. It defines minimal interoperability requirements but was not designed for real time access and process interoperability. It is difficult to apply in B2B data sharing settings where time is a critical factor. Additional interventions where the necessary infrastructure and interoperability requirements are mandated would be necessary to achieve real-time access.

There are no legal provisions for **portability for non-personal machine data in IoT** settings. There is no law that establishes an unequivocal legal link between the machine or party that collects the data and the party that would have the right to access the data. Machine data are often co-generated; several parties can claim access rights. The idea to introduce data ownership rights for non-personal data as a way to establish exclusive rights has somewhat faded and is now replaced by access rights. However, re-labelling the problem does not change the nature of the issue: how to identify the parties that could claim control and access rights among several claimants. The allocation of rights that maximises social welfare may be perceived as unfair and raise equity concerns for disadvantaged parties.

Portability has several economic effects. It may **increase the volume of B2B data sharing and reduce the market power of a data holder**. Re-users that want to collect data via portability incur costs to incentivise portability rights holders to initiate a data porting request with their data holder. As a result, economies of scope in data aggregation are often weaker for re-users than for the original data holder. That may reduce the efficiency of re-use-based services and put them at a competitive disadvantage compared to the original data holder. Conversely, data portability may strengthen economies of scope in data aggregation when a major market player manages to leverage portability rights to collect data from smaller and fragmented service providers and aggregate them in a larger data pool that generates efficiencies in service production compared to the original holders of fragmented datasets. In such cases data aggregation **may lead to increased market concentration** and efficiency losses because of reduced competition.

With regard the second scenario, **B2B data access** differs from portability because the data originator or holder has no role to play in the data sharing decisions. Sharing can be initiated by a designated access rights holder. The **problem is to identify these rights holders in complex machine data and IoT settings with several stakeholders**. For non-personal machine-generated data there is not always an obvious "natural" party that can claim data access rights. The machine manufacturer is in a privileged position to design the machine in such a way that he has exclusive control over the data. In that case, it may be better to shift access rights away from machine owners/operators and data originators and assign them directly to specific groups of data re-users.

This raises questions about the **decision criteria to grant such rights**. Maximizing social welfare for society as a whole, and the failure of the market to do so, is a good starting point. For example, when data pooling is unlikely to be produced by the market because data holders have no incentives to contribute to this pool unless they can directly benefit from the insights and services produced by the pooled data. Yet, there may be winners and losers from data pooling. This complicates social welfare considerations with static equity and welfare-redistribution questions and with dynamic innovation questions.

Introduction

The digital transformation generates ever larger volumes of data. Data are not only collected, they are also shared and traded between parties. As shown by Everis (2018) and IDC and Lisbon Council (2019), markets for data are growing rapidly in many domains, from advertising to financial markets, maps and navigation services and many other applications. The potential benefits of data sharing and re-use in many applications are acknowledged (Ctrl-Shift, 2018; Fingleton and ODI, 2019; Kramer, Schnurr and Broughton-Micova, 2020; OECD, 2015 and 2019). Firms are encouraged to engage in data sharing, making their data available to other users. Yet, while data sharing can stimulate innovation and ensure a level playing field in competition between firms, it can also entail commercial risks and might weaken firms' incentives to collect data. Finding an appropriate balance between opening access to data and keeping them private for firms and individuals is one of the most important and complex tasks for modern digital societies (Palfrey and Grasser, 2012). Data sharing is not a policy objective in itself but a tool to promote economic and social welfare for societies.

The European Commission's Data Strategy (2020, p 7) starts from the premise that there is not enough data sharing and much data remain locked up and are not available for innovative re-use. It identifies several obstacles to data sharing, including legal fragmentation between EU member states, lack of trust and imbalances in market power between businesses, lack of data interoperability and common storage spaces, and tools for empowering individuals to exercise their data rights. The Strategy seeks to promote business-to-business (B2B) and business-to-government (B2G) data sharing with a wide range of legislative and other initiatives, including standards to promote interoperability, the creation of data pools, improving portability rights and even changes in competition tools with respect to data. It acknowledges the need to protect and empower individuals and firms to exercise their data rights.

The purpose of this report is to inform the Data Strategy with an examination of the causes of obstacles to data sharing. What keeps businesses from sharing or trading more of their data with other businesses or governments? For this purpose, this report uses the well-known tools of legal and economic thinking about market failures. Data markets, and related data-driven services markets, fail when they underperform and do not produce the social welfare for all that they could potentially produce because the behaviour and incentives of private operators locks markets into an inefficient situation. According to the European Commission Better Regulation Guidelines (2017b), public policy intervention may be justified when there are market failures, regulatory failures, equity concerns or behavioural biases that result in inefficient outcomes that are not in the public interest and hold back the overall welfare of society. Policy interventions can take various forms, including the use of competition enforcement tools to create a level playing field in data-driven markets and regulatory initiatives that facilitate data sharing, or make it mandatory.

Of course, there might be other justifications for intervention in data markets than market failures. Data interventions can be justified from a fundamental rights perspective, as in the case of rights for data subjects in the EU General Data Protection Regulation (GDPR). As we focus on B2B data sharing, we are dealing either with non-personal data or anonymized personal data, or data subjects have given their consent for

further sharing. While consented data retain some inalienable fundamental rights characteristics, non-personal machine data have no fundamental rights attached. Market failures and fundamental rights can however interact. For example, the Data Strategy mentions the promotion of user empowerment and self-determination by means of personal data portability, a personal data right defined in the GDPR. The portability right can affect market outcomes, as we will argue in this report.

This paper is structured as follows. Section 1 discusses the specific economic characteristics of data that make data sharing an attractive economic proposition. We also discuss the interaction between data markets and markets for ordinary goods and services. Section 2 explores different causes of data market failures in B2B data interactions between two or more firms. Section 3 examines how some types of data market failures may be overcome with third-party intermediaries, data pooling and other technological solutions. Section 4 shifts to data market failures that require government intervention by means of competition instruments and regulatory policy. Finally, Section 5 summarises the findings and suggests some ways forwards for policy makers to decide on the most appropriate actions to take, depending on the type and source of market failures in data markets and data-driven services markets.

1. The economic characteristics of data and data markets

1.1. The economic characteristics of data

Before we dive into data markets and sharing, it is important to explain a few key economic characteristics of data that set them apart from ordinary goods and services.

A crucial economic characteristic of data is non-rivalry[2]: many parties can use the same dataset for a variety of purposes without functional loss to the original data collector (OECD, 2016; Jones and Tonetti, 2019). 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 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. Non-rivalry is the fundamental driver of economic welfare gains in data sharing or re-use: if one firm collects data that can be used for many purposes[3], society would benefit if other firms could access and use these data. It would result in cost savings for society (no need to re-collect the same data by other means, even if feasible). It would enable the production of new and innovative data services that the original data collector had not envisaged. The primary data collection effort is a sunk cost that can be amortised across a large number of uses, rather than remaining confined to a single user.

An economic interpretation of the benefits of non-rivalry at the firm level revolves around the concept of **economies of scope that occur in joint production and (re-)use of the same product or asset to produce several outputs** (Teece, 1980, 1982; Panzar and Willig, 1981). For example, a car manufacturer can re-use the same chassis and engines to produce different car models. In this case the advantages of re-use pertain to the fixed cost of creating the design for a chassis or engine. Re-use of the same non-rival design entails zero marginal re-design costs. However, there is a positive marginal cost for the production of additional rival chassis and engines. More generally, non-rival immaterial products, such as proprietary knowledge and digital data, have quasi-zero marginal re-use costs because it does not involve re-producing a physical good, only copying an electronic data file.

Besides re-use, there is a second source of economic efficiency gains in the use of data: **economies of scope in data aggregation**. When two datasets are complementary, more insights and economic value can be extracted from merging them, compared to keeping them in separate data silos. This insight can be traced to the economics of learning and division of labour. Rosen (1983) 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 and learning one skill decreases the cost of learning another, then there are

[2] Kramer, Senellart and De Streel (2020) distinguish between the non-rival nature of data and rivalry in the means of data collection because key data-collecting services (such as search, or social networking) are highly concentrated around a few firms. These data are not ubiquitously available. Both perceptions provide economic arguments for data sharing.

[3] To the extent that a single dataset can be used for several purposes, data could be considered as a general-purpose technology. However, this re-purposing may not reach the same degree of pervasiveness and cross-sectoral applications as envisaged by Bresnahan and Trajtenberg (1995).

economies of scope in learning both skills, provided that the benefits from interaction exceed the additional learning costs. Similarly, when two datasets are complementary and not entirely separable, applying data analytics (i.e., learning techniques) 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 an extended dataset is small. For example, a targeted advertising service is likely to become more efficient in terms of click-through rates and revenue when it has access to consumer data that combine a variety of sources including web browsing, financial transaction data, mobility data and social media messages. Targeted advertising strategies can be derived from each of these datasets separately but the combined dataset will yield more insights into consumer preferences.

The **benefits of economies of scope in data sharing and re-use by other firms may also have a cost side**. Palfrey and Grasser (2012) warned that all digital data can, in principle, be made interoperable and shareable to the benefit of society. However, firms and persons may also suffer from sharing. 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 without functional losses, sharing may entail economic losses for the original data holder. Firms and persons will **trade off the benefits they expect to receive from sharing (trading) their data against the costs they may incur from doing so**. Benefits will stimulate private markets for data sharing while costs will put a limit on data transactions.

The question for policy makers is **whether these private markets maximise the joint social welfare** of data originators, holders and users. If not, then there is a market failure that may require policy intervention. Policy intervention should not seek to maximize data sharing because data sharing is not an objective in its own right. It is a means to achieve higher social welfare in society. From a market failure perspective, 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.

Economies of scope should be distinguished from **economies of scale in data**.[4] A useful way to illustrate this 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. Statistical analysis can be applied to the dataset, for example to use them for prediction purposes. Economies of scale refer to increased prediction accuracy due to an increase in the number of rows. Economies of scope refer to increased prediction accuracy due to an increase in the number of columns or explanatory variables. Adding more columns (variables) is not helpful when they are highly correlated or when they are not related at all.

Ordinary goods are excludable by nature. If one person has it, another cannot have it at the same time. **Data are not excludable by nature because they are non-rival**. Two persons can have the same data at the same time. In order to sell data, a data holder must have exclusive control of the data. If more parties hold the same dataset,

[4] Also Lerner (2014).

competition will drive prices down to their marginal cost of reproduction, often close to zero. **Excludability can be achieved by technical means**. The *de facto* data holder can apply technical protection measures to ensure his exclusive control and access to the data. Excludability can also be **achieved by legal means**. The data holder can negotiate a bilateral contract with a buyer or user that specifies data access and use conditions. Contracts can be enforced in courts. In some cases, the law grants *erga omnes* exclusive rights. For example, the EU General Data Protection Regulation (GDPR)[5] grants natural persons some exclusive rights to their personal data. The EU Database Directive[6] grants a conditional *sui generis* right to makers of databases. These exclusive rights strengthen the economic position of the rights holders in data markets – very similar to the position of IP rights holders in patent and copyright markets. The law can also grant a set of defensive rights, notably if data can be considered as "trade secrets", as defined in the EU Trade Secrets Directive[7].

Data are usually not a final consumer product. They are an **intermediary input** into the production of a service. For example, 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. Consequently, data sharing implies a relationship between an upstream data holder and a downstream data-driven service producer who trade or share data between them – unless the two roles are combined in a single firm.

Data sharing is a label that may cover different economic modalities: sharing for free, trading for a monetary compensation or in exchange for other data, direct sharing of a dataset or indirect sharing of a data-based service only. The latter implies that there is no exchange of data between two parties; there is only an exchange of a service based on data. For example, online advertising platforms like Google do not deliver detailed consumer data to the advertiser but only an advertising service, based on broad targeting criteria. Data have no value on their own. They only become valuable to the extent that parties can use them to leverage their position in data-driven services markets.

A peculiar characteristic of data is their **social value**. Economies of scope in aggregation are a first source of social dimension value. Two owners of separate but complementary datasets can only achieve a higher value from their data if they collaborate and pool the two sets. A second source of social value comes from economies of scale. Once a sufficiently large sample of behavioural observations has been compiled to produce robust predictions, that can be used to predict the behaviour of agents outside the sample[8]. These social externalities imply an inherent market failure in exclusive private control over data. The party that does provide the data is not necessarily 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. An intermediary agent may be required to realize

[5] Regulation 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 (General Data Protection Regulation) [2016] OJ L 119/1.

[6] Directive 96/9 of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L 77/20.

[7] Directive 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure [2016] OJ L 157/1.

[8] Bergemann, Bonatti, and Gan (2020); Acemoglu et al. (2019).

the social externalities from data pooling and turn them into benefits that pay for the coordination costs and incentivise individuals to participate in the pool.

The social value of data brings us to online **digital platforms as intermediaries that can realize the benefits from economies of scale and scope in data aggregation**. Data played no role in the first generation of economic models of platforms or multi-sided markets[9] that were extension of the economics of infrastructure networks. Recent platform models have broadened the definition of platforms to firms that bring economic agents together and actively promote social externalities between them[10]. They focus on data-driven network effects in platforms that increase the social value of data[11]. For example, platforms can create a searchable catalogue of products or a directory of users as a first step in generating that social value. For more efficient matching in complex search engines, platforms collect detailed data on buyer preferences and product characteristics. For example, Netflix can improve its film title search engine when it learns more about user preferences and film characteristics[12]. Platforms are in a unique position as third-party data aggregators to realize economies of scale and scope in data aggregation across many users. Individual users cannot realize these social benefits on their own.

The social value of data raises the question which **welfare measure should we use to assess if there is a gap between private and social welfare**? The mainstream view in competition law is to use consumer welfare as the reference point.[13] Overall however, economics normally uses a wider social welfare measure that comprises of **consumer and producer welfare**, or the combined welfare of all stakeholder groups in society. Some competition lawyers would accept this wider view. These two measures can easily lead to contradictory conclusions, for example when price discrimination shifts consumer surplus to producers or between consumer groups. Social welfare measures also have their problems[14]. 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 welfare improvement whereby some parties gain and others lose[15]. This raises equity concerns.

Another discrepancy in welfare measures may occur when we compare **static competition scenarios with a dynamic innovation scenario**. Static scenarios examine the welfare and re-distribution effects of a given market structure and pricing strategy, for a given technology. Dynamic scenarios include the impact of future

[9] Caillaud and Jullien (2003); Parker and Van Alstyne (2005); Rochet and Tirole (2003); Rochet and Tirole (2006).

[10] Franck and Peitz (2019). This definition does avoid the problem of setting a minimum number of market sides; one is enough.

[11] Data-driven network effects were first analysed by Prüfer and Schotmüller (2017).

[12] Iansati and Lakhani (2020: ch. 6).

[13] See Motta (2004).

[14] The notion of social welfare can of course be stretched beyond traditional market failures in economics and include broader societal issues such as equity and income distribution, the protection of vulnerable groups and minorities, and the protection of cultures and political systems. For example, the protection of liberal democracies against fake news and online disinformation or the right to self-determination of citizens in the presence of artificial intelligence.

[15] Economics distinguishes between strictly Pareto-improving welfare measures whereby no agent loses welfare. A less stringent welfare improvement criterion is the Kaldor-Hicks criterion whereby agents that lose some welfare could by hypothetically compensated by the gains that other agents make. It does not imply that the compensation payments will actually be made.

innovations on the longer-term welfare of citizens. This is especially important when re-use and aggregation of datasets can trigger significant innovations.

1.2. The interaction between data collection, use and services markets

Competition lawyers and economists always ask: what is the relevant market? Since data derive their value from use in services markets, we have to look at least at two markets: **data collection and use markets**. Firms collect data in a primary market. Data originators require incentives in order to share their data with a collecting firm. That firm stores and processes the data for onward direct or indirect sales to data-using firms in a secondary market. That, in turn, will affect dynamics in data-driven services markets[16]. These markets interact. The willingness of data originators to share data with collectors will not only depend on sharing conditions in the primary market but also on subsequent use of the data by the collecting firm in the secondary market. For example, the willingness of consumers to share their data with a website will depend on the quality of services offered by that website to the consumer as well as on how the website will subsequently use the data for advertising or other purposes.

Some recent data economics papers have started to look at this **two-sided dimension of data markets**. Jones and Tonetti (2019) use a theoretical **macro-economic growth model to illustrate the social welfare gains from data sharing**. Firms use data to increase the efficiency of production and, because of non-rivalry in the use of data, to increase the variety of goods and services in the economy. Access to a larger data pool will boost the productivity of firms and the number of innovations[17], and thereby contribute to economic growth and consumer welfare. However, individuals may reduce the amount of data that they share because of privacy concerns. Similarly, firms consider the impact of sharing their data on competition and innovation – the emergence of close substitute products – in their markets. Data hoarding increases the private welfare of persons and firms but slows down innovation and economic growth.

Jones and Tonetti (2019) consider several policy scenarios: the optimal degree of sharing determined by a benevolent social planner, private sharing decisions by persons and firms that get ownership or control rights over secondary use of their data in order to alleviate their concerns, and completely outlawing data sharing. (i) If firms own the decision rights they will be more willing to selectively share the data they hold. However, the volume of data they receive from consumers is substantially diminished because consumers fear for their privacy since they have no rights to control the use of their personal data in this scenario. (ii) If consumers have control rights they share a larger volume of data because they feel more in control. This boosts economic growth. They show that **allocating data control rights to private persons is superior to allocating these rights to firms that trade data**.

[16] Of course, there can be vertical integration between these markets. A firm may collect consumer data as a by-product of its ordinary transactions and use the data in another service market in which it is active.

[17] This is an endogenous growth model. Product variety is the result of increased growth and investment that follow in the wake of data availability. It is not directly the result of increased data availability and the innovation that this might produce.

Acemoglu et al (2019) argue that **externalities in personal data collection create a market failure and diminish the value of personal data in the primary market**. In the age of artificial intelligence and machine learning, data collected on the behaviour of one set of consumers has predictive value for the behaviour of other consumers (Agarwal et al., 2018). This is essentially an "economies of scope in data aggregation" argument. Once a firm has accumulated a critical mass of consumer data, the additional insights obtained from adding another consumer's personal data are small, compared to what can be learned from data already collected about persons with a similar profile. This reduces the marginal value of a single person's dataset and diminishes incentives for consumers to protect their privacy. That, in turn, increases the supply and decreases the market value of personal data.

This **could explain the privacy paradox** (Acquisti et al, 2016): consumers value their privacy but do not invest in protecting it because they understand the low value-added of their personal data and the futility of these efforts in the presence of strong spill-overs from other consumers' data. Consumers may not be that sophisticated in their thinking and still invest in privacy protection. However, that investment in itself may have a signal value that can be exploited against their interests (Dengler and Prüfer, 2018). Moreover, re-use of personal data has ambiguous effects on consumer welfare. 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 aims to be more persuasive than informative and drives consumers away from their original preferences.

While none of these papers presents a two-sided or multi-sided market model for data, the multi-sided nature of data markets is implicitly or explicitly present in all these models. We therefore conclude from these papers that **conditions in the primary data collection market have an important effect on the availability of data in the secondary market for data re-use, and on data-driven services markets**. **Conversely, re-use conditions will affect the operations of the primary data collection market.** Data re-use in the secondary firm-to-firm or business-to-business (B2B) market can therefore not be considered in isolation from the primary market.

The above-discussed papers focus on the allocation of data ownership and control rights to firms and natural persons. However, in practice, **there are no legal ownership rights on data**[18], in the EU or elsewhere (Duch-Brown et al., 2017). The EU GDPR grants natural persons some inalienable control rights over their personal data. However, it does not recognize tradable ownership rights in personal data[19] because data protection is a non-alienable human right. This sparked a debate on the need to introduce such ownership rights, at least for non-personal data (European Commission, 2017a). This idea has gained very little policy traction so far, mainly because of complications regarding the allocation of such rights. Should data originators, collectors or processors have such rights? Recent policy reports have shifted the debate to access

[18] Apart from a limited ownership right on databases, recognised in the EU Database Directive. In addition, data can be protected by copyright and/or as a trade secret if it meets the relevant conditions.

[19] Though the EU Digital Services Content Directive has taken some steps towards recognizing that personal data can be tradable: Directive 2019/770 of the European Parliament and of the Council of 20 May 2019 on certain aspects concerning contracts for the supply of digital content and digital services [2019] OJ L 136/1.

rights to existing datasets (European Commission 2018a; Drexl, 2018). However, the focus on access rights does not fundamentally change the debate. The question still remains who should have such rights, and under what conditions?

The EU GDPR allocates personal data rights to the data subject, a natural person who is the source of the data. In the case of non-personal machine data generated in industrial processes where many parties intervene, the allocation of rights is not self-evident and may trigger substantial shifts in added value in the production process. In Internet-of-things settings, machine data often end up under the *de facto* exclusive control of one party because sensors and machines are designed to achieve that outcome.

In the next section, we explore these questions from the perspective of market failures as a benchmark for possible intervention in private data markets.

2. Private markets for B2B data sharing and market failures

In this section we focus on B2B data sharing or the market for data re-use and aggregation.[20] We consider the primary data collection market as exogenous and examine possible market failures in the secondary B2B market that might justify a policy intervention. A number of recent competition and data policy reports (Crémer et al, 2019; Furman et al, 2019; Scott-Morton et al, 2019) discuss market failure in secondary data markets from the perspective of monopolistic behaviour and competition policy. Most of that debate is situated in the context of very large online platforms having a strong position in data markets and related data-driven services markets. They may leverage data-driven network effects (Prüfer and Schottmüller, 2017) and economies of scope in re-use to strengthen their position in adjacent markets or exclude others from entering the market. Here we will also start from monopolistic data markets and competition-related market failures (Section 2.1). However, we expand the debate and look at other causes of market failures too including externalities, lack of incentives for the production of non-excludable (public) goods, missing markets, and imperfect and asymmetric information (Section 2.2).

2.1. Monopolistic data markets

We start from the perspective of a private firm (Firm 1) that collects a relatively scarce dataset (D1) for which there are no close substitutes. The firm consequently benefits from a monopolistic market position. We assume that Firm 1 uses D1 for the production of a service S1 so that the production of D1 and S1 are vertically integrated[21]. D1 may be collected independently of S1, or it may be a by-product of producing S1. For instance, consumer data collected while providing an e-commerce or social media service[22]. We assume that D1 can be re-used to produce another service S2 by Firm 1 or by another Firm 2 (economies of scope in re-use), or a service S3 that requires aggregation of D1 with another dataset D2 owned by Firm 2 (economies of scope in aggregation). In order for society to realize the social welfare gains from the potential economies of scope that D1 could generate, S2 has to be produced by either Firm 1 or Firm 2, and the production of S3 requires coordination or collective action between Firm 1 and Firm 2.

We examine the obstacles that might impede the realization of these economies of scope. We start from the incentives that Firm 1 faces to maximize profits from D1, in three steps:

[20] See also tow studies done for the European Commission on issues raised in practice by data sharing: Deloitte et al. (2018) and VVA (2017).

[21] D1 and S1 can also be carried out by two firms. That does not fundamentally change the reasoning in this section.

[22] The "by-product" assumption is often invoked to justify open and free access to Firm 1's dataset D1 (OECD, 2016; European Commission, 2018), the argument being that it would not have any negative economic impact on F1; sales revenue of S1 would continue. This assumption ignores that F1 would no longer have revenue from selling D1 to other firms. It also ignores that D1 sales revenue may have an incentive effect on collecting more data, for instance by lowering the price of S1, especially in a two-sided market setting. Access to D1 by other firms may reveal F1's commercial strategies that could be used by other firms to leverage their market share in S1. It assumes that costs are separable and identifiable for main and by-products. This is not necessarily the case. Granting mandatory free access rights to D1 eliminates F1's data monopoly rent and dissipates it towards downstream data users. This is sometimes justified as a measure to promote data-driven innovation. It re-allocates rents from static monopolistic equilibrium to dynamic innovation equilibrium.

- The impact of the substitution effect between S1 and S2 on the re-use of D1;
- Data trade or vertical integration between Firm 1 and Firm 2 for the production of S2;
- The internalisation of externalities from aggregation of D1 and D2, or the benefits from coordination between Firm 1 and Firm 2.

(a) The substitution effect in secondary services markets

A first question is whether Firm 1, that produces S1, has an incentive to also produce S2 or not. The definition of non-rivalry suggests that the original use is not affected by re-use for another purpose. While re-use may not functionally affect original use for S1, it may however have an economic impact on the original data collector[23]. The second use can complement, substitute or be neutral with respect to the first.

If S2 is a competing substitute for S1[24], **Firm 1 will try to prevent the production of S2 because it undermines the market for its own service S1** (Zhu et al., 2008, Jones and Tonetti, 2019)[25]. For example, car manufacturers will be reluctant to share car maintenance data with independent maintenance service providers that compete with their own network of official dealers (Martens and Müller-Langer, 2020). Another example of competing services is data collected by taxis and public transport service providers to manage their user services that could be accessed by competing mobility service providers such as e-scooter and ride hailing apps, for instance in the context of urban mobility services platforms. They could be used for commercial strategies that seek to reduce the number of customers and revenue of the original data collectors (Carballa-Smichowski, 2018)[26].

If S2 is a complement to S1, **Firm 1 has an incentive to facilitate the production of S2 because it will increase the sales of S1**. For example, car insurance and navigation services complement car sales. Car manufacturers have an incentive to use data to reduce the costs and improve the quality of these aftermarket services because that increases car sales.

---

[23] In that sense, data are club goods: more users of the data may initially increase the value but too much use may negatively affect the value. While they are non-rival, economic value maximisation requires some degree of excludability. Unlimited sharing of non-rival data is not optimal (Bergemann and Bonatti, 2018; Palfrey and Grasser, 2012).

[24] Two services are substitutes when a decrease in the price of one decreases demand for the other. They are complements when a decrease in the price of one increases demand for the other.

[25] In the context of the EU database directive and its possible equivalent into US legislation, Zhu et al. (2008) present an economic model that explains the conditions that should be attached to re-use of a database. They discuss three factors that have played an important role: substantial expenditure for the creation of an original database, the extent of functional equivalence between the reused data and the original, and injury for the original creator. They translate this into an economic model with three key variables: fixed investment costs for the creator, substitutability versus complementarity between the original and reused data, and impact on the revenue of the creator and re-user. If the re-use is a complement rather than a substitute to the original, the two parties will not compete in the same market and revenue for the creator will not be affected.

[26] This is a standard trade-off in vertical foreclosure settings. Upstream monopolists that aim to foreclose a downstream market need to take into account possible positive or negative effects on their upstream market. Potential participants in such data sharing platforms need to consider the balance between positive market expansion and negative competition effects. The data supplier will have an incentive to refuse data supply unless he receives sufficient compensation for potential revenue losses. Both sides may find it individually optimal to only partly meet the demands of the other side. This may result in an overall suboptimal outcome that reduces the functionality of the mobility service and the social welfare benefits from data.

Alternatively, S2 can be neutral with respect to S1, neither a complement nor a substitute. For example, mobility data generated by mobile phone service operators can be used to enhance city traffic management, a service that does not compete nor complement the original use. Selling mobility data for traffic management generates purely additional revenue for mobile phone operators.

In conclusion, **economies of scope in re-use may not be realised if S1 and S2 are close substitutes**. This may entail welfare losses for society, especially when substitutes would increase competition in downstream service markets. Examples include data-driven competition in car maintenance, in payment services and in energy distribution markets.

(b) Vertical integration

A second question for Firm 1 is whether to produce S2 in-house or sell access to D1 to Firm 2 to produce S2. This is a vertical integration question. The answer depends on which option is the most profitable for Firm 1. Since the marginal cost of re-using D1 for the production of S2 is close to zero, profitability will be determined by the monopoly price that Firm 1 can extract from Firm 2 for access to D1. (i) ***If Firm 2 can obtain a substitute dataset D2 for the production of S2***, pricing of D1 would have to consider the cost of alternative D2. D2 may be an imperfect substitute that produces a lower quality service S2 that fetches a lower market price. For example, producers of car insurance and navigation services can switch to alternative providers of car navigation data, such as mobile phone operators, to produce a competing service S2. Still, the car manufacturer may decide that his own service S1 can compete with S2 and that this option is more profitable than selling the dataset D1. Market conditions in the data input market as well as the services market will affect pricing strategies for Firm 1's dataset D1. (ii) ***If there is no alternative dataset D2***, Firm 1 has a monopoly on an "essential facility" for the production of S2 and may price D1 in a monopolistic way. This can lead to market distortions discussed extensively in recent competition policy reports (Crémer et al, 2019; Furman et al, 2019; Scott-Morton et al, 2019)[27].

A profit-maximizing Firm 1 will ration data sales in order to maximize the scarcity value of his data. (i) ***If price discrimination between buyers is not possible***, Firm 1 may sign an exclusive deal with Firm 2 and foreclose the market for other firms that wish to access D1 (Montes et al, 2017). This reduces competition in the market for S1 and reduces economies of scope in the re-use of D1. It leads to social deadweight losses in both data and services markets. (ii) ***If price discrimination is feasible***, Firm 1 may extract all surplus from downstream users of D1, possibly leading to equity concerns in the distribution of welfare. There is a growing volume of economic research on a variety of data sales and price discrimination strategies to maximize revenue for a monopolistic data holder (Bergemann and Bonatti, 2018).

In an extreme case of foreclosure, Firm 1 merges with Firm 2 to produce a joint service S2, rather than trade the data. This changes the way in which Firm 1 monetizes the value of D1, from a sales contract to in-house processing in the merged

[27] For a more theoretical model of data-driven vertical integration see de Cornière and Taylor (2020).

firm. Trade may not be possible because of regulatory constraint, for example because of purpose limitations and re-sale constraints imposed on personal data under the EU GDPR. Vertical integration may be a way to overcome these regulatory constraints and aggregate personal data from various sources. Vertical integration has pricing strategy advantages for the data monopolist and for society. It avoids double marginalization in the sequential pricing of D1 and S2 and could lead to more efficient market outcomes for consumers of S2. De Cornière and Taylor (2020) study the impact of a merger on the primary market - the incentives to invest in collecting D1 – and on competition in the secondary market where D1 is re-used for the production of S2. They find that, if data trade is not possible, the merger increases consumer welfare because it increases data-driven competition in the market for S2. If data trade is possible, the merger reduces the incentive to collect more data D1 and thereby diminishes competition in the S2 market because the quality and/or quantity of data available for use in S2 is degraded compared to a situation without data trade.

There may also be **intermediate solutions between selling the data to Firm 2 and merging with Firm 2**. For example, large online consumer platforms may grant temporary and limited data access rights to a potential innovator Firm 2 that claims it can use D1 to produce S2. Firm 2 can get a data carve-out inside Firm 1's server system – to avoid data leakage risks – and experiment with the data in the market for S2 for a limited time period (Parker and Van Alstyne, 2017). If the innovation is successful, the two firms may apply a pre-agreed protocol to split the benefits; if it is not successful, data access is simply closed. The platform may also allow a temporary transfer of data to Firm 2 to experiment with innovative uses of the data in the development of new services. In the absence of a prior protocol, the data holder may cut off data access for the innovator and free ride on the innovation without compensation[28]. However, risks of data leakage and misuse may be high, as the Facebook and Cambridge Analytica case showed.

(c) The need for complementary inputs

A third question for Firm 1 occurs when the production of S2 requires complementary inputs that it does not have. ***If the market for these complementary inputs is competitive***, it can buy them to ensure in-house production of S2. Even so, there may be fixed costs and economies of scale in these inputs as well that lead to market imperfections because acquiring the inputs becomes prohibitively costly for Firm 1. In that case, Firm 1 is better off selling D1 to Firm 2 that already has these complementary inputs. Fixed costs may thus work both ways. Fixed costs in the collection of data may give Firm 1 an advantage in the production of S2. But fixed cost in complementary resources may shift these advantages to other firms. This will affect switching between internal production of S2 and trading data D1 with another firm for the production of S2. Bourreau and de Streel (2019) go back to the economic literature on "conglomerates" to show how economies of scope in traditional firms may have

[28] The PeopleBrowsr v. Twitter (Superior Court of the State of California, *PeopleBrowsr, Inc. et al. v. Twitter, Inc. (PeopleBrowsr)*, No. C-12-6120 EMC, 2013 WL 843032, N. D. Cal., 6 March 2013) and HiQ v. LinkedIn (United States District Court, Northern District of California, *hiQ Labs, Inc. v. LinkedIn Corporation*, No. 17-cv-03301-EMC, 14 August 2017) cases illustrate this point.

contributed to firms' expansion of activities in many areas. Economies of scope in data re-use may lead to conglomerates in the digital economy.

If the market for the complementary input is monopolistic, a data monopolist and a resource monopolist will have to work out an agreement to share their respective production factors in order to produce S2. This is generally known in economics as an *"anti-commons" problem* (Buchanan and Yoon, 2000; Schultze et al., 2002). Two parties have exclusive rights over resources that need to be combined in order to produce a service that is in their common interest. They need to cooperate and negotiate the allocation of costs and benefits of combining the resources for the production of S2. This leads to strategic behaviour whereby owners try to internalize benefits for themselves and externalize costs to others, and results in Nash bargaining that leads to a Pareto-inferior suboptimal solution because all holders of exclusion rights aim to maximize their own profits and set a monopolistic price. As a result, the combined price is higher than the optimal price and the produced quantity is less than under full monopoly by a single party. The uncoordinated exercise of exclusion rights leads to under-utilization of data. S2 will not materialize, or only in an inferior quality and quantity. Unless there are market-based solutions to overcome this coordination failure, there may be a need for policy intervention.

(d) Economies of scope in data aggregation

A specific case occurs when the production of a new service S3 requires complementary inputs from D1 and another monopolistic dataset D2 owned by Firm 2. This is a typical case of economies of scope in data aggregation. Firm 1 and 2 need to come to an agreement as a pre-requisite for the production of a joint service S3. The resulting coordination problems have been amply discussed in the common and anti-commons literature. There are many examples that show that private firms and markets are often able to overcome the coordination problems to achieve economies of scope in data aggregation, sometimes with the help of a third-party intermediary. For instance, car manufacturers only have access to navigation data from cars from their own brand, not from other brands, which makes it difficult to produce accurate traffic congestion maps. Several manufacturers decided to collaborate to share car navigation data in a joint navigation service HERE, thereby improving the quality of navigation services (Martens and Müller, 2020). Another example comes from health services where Google's Deep Mind negotiated access to aggregated consumer data from UK health service providers. Although the case is controversial, it is likely that the application of data-intensive artificial intelligence techniques to the aggregated data may contribute to discovering new disease patterns and medical treatments.

While traditional economies of scope in re-use constitute an argument in favour of wider access and diffusion of non-rival datasets across many agents, **economies of scope in aggregation underline the economic benefits of data concentration in large pools**, as underlined in the European Commission Data Strategy (2020, p 6). The two are not mutually exclusive. Accumulating data in large pools requires combining smaller datasets. Large data pools do not necessarily imply monopolisation when there are several overlapping and competing data pools. Nevertheless, economies of scope in data aggregation constitute an incentive for digital conglomeration (Bourreau and de Streel,

2019). It may also strengthen market dominance because it increases entry barriers for new firms and diminishes incentives for innovation (Kramer, Schnurr and Broughton-Micova, 2020; Prüfer and Schottmüller, 2017). Once a firm has built a strong data position in one domain, the marginal costs of expanding into an adjacent complementary data domain are lower than for *de novo* entrants in that domain or incumbents who only cover that specific domain. Data-driven indirect network effects may be too strong for a new entrant to compete with. McNamee (2019) illustrates how Google expanded its search data to adjacent location and mobility data to create maps and navigation services, and Facebook appended its social media data with browser cookie and mobility data. The value of the aggregated sets exceeds the sum of values of the separate sets.

Schultze et al. (2002) offer some insights into **how segmented data markets may still achieve collaboration**. They explore how differences in supply side cost structures and expected net benefits between the participants in an anti-commons game affect outcomes. Asymmetric market power and cost structures may lead to more data sharing than under the Nash equilibrium predicted by Buchanan and Yoon (2000). A dominant data player can offer a share in the additional value generated by economies of scope from combining two datasets that exceeds the value that a small player can achieve from his own dataset. However, the distribution of added-value between the dominant and smaller player may be very unequal. Another solution is the **introduction of a third-party intermediary** that ensures the enforcement of commitments agreed between the bargaining parties in order to overcome the Prisoner's Dilemma situation.

Yet, there may be situations where inefficient private bargaining or the lack of incentives to come to an agreement might justify government intervention to facilitate data access and pooling in order to achieve the welfare-enhancing benefits of economies of scope in data aggregation. This might be at the expense of a shift in the welfare distribution in society: some may gain and others may lose in order to enable the overall gain. If we accept this position, then forcing data sharing on some agents, for the benefit of the group becomes a possible policy option.

Economies of scope in data aggregation can have an ambiguous economic effects (Lundqvist, 2018; Richter and Slowinski, 2019). It may be an anti-competitive force in the data economy if it involves collusion through the exchange of commercially sensitive information among competitors. It may turn into abuse of dominance when the aggregated dataset is used to anticompetitively leverage market power to the detriment of the aggregators in downstream services markets. That may be subject to scrutiny by competition authorities under Articles 101 and 102 TFEU. At the same time, economies of scope in data aggregation can generate productivity and social welfare gains, and innovation benefits, for society. For example, pooling detailed medical data from many patients and health service providers can increase the productivity of medical research and ultimately benefit society as a whole. Moreover, there can also be re-distributional effects. Some consumers and health service providers may experience negative effects from the insights produced by data aggregation when the results single them out for discriminatory treatment. The results may not be welfare enhancing for all agents[29].

[29] Economics distinguishes between strictly *Pareto*-improving welfare measures whereby no agent loses welfare. A less stringent welfare improvement criterion is the *Kaldor-Hicks* criterion whereby agents that lose some welfare could by hypothetically compensated by the gains that other agents make. It does not imply that the compensation payments will actually be made.

2.2. Other causes of data market failures

In Section 2.1 we discussed how the market power of the data holder may prevent, under some conditions, the realisation of economies of scope in re-use and aggregation of data. In this section we bring together other sources of data market failures, such externalities, missing incentives for the production of public goods, incomplete contracts and risks, transaction costs and missing markets, and asymmetric and imperfect information that distort decision making. These causes are not neatly separated; they may overlap and interfere with each other.

(a) Externalities

Externalities emerge when a decision by one party creates a spill-over of costs and benefits to other parties that are not involved in the decision-making. When data collected by one party can be freely accessed by others, it results in free riding on the efforts of the original collector. This **generates positive externality benefits for re-users and aggregators and may be social welfare-enhancing for all**. In a normal data trading relationship, the original data collector would expect to appropriate at least some of these benefits, for instance by selling the data for a price rather than allowing free access. Monetisation (partly) internalizes the externalities. The original data collector may depend on monetisation revenue to finance his data collection efforts. Even if data collection does not depend on revenue from re-use, firms that collect data in the course of their regular business activities may want to explore new ways of monetizing the data. For example, mobile phone operators can sell user location data to traffic authorities, car navigation services and advertisers. Monetisation of re-use re-distributes benefits between collectors and re-users and internalizes the externalities, giving additional incentives to data collectors.

However, some data externalities may be hard to capture and monetize, for example in health services. Data aggregation across private health service providers could create a very useful database with strong economies of scope in aggregation for the purpose of analysis of disease patterns, to improve the efficiency of health services and develop new cures, medicines, prescription policies and public health policies. Yet, private data collectors have no incentive to contribute their data for the production of these services because they cannot monetise these benefits that spill over to other parties. Data aggregation may even generate negative spill-overs for the original data collector. For example, the aggregated data could be used for comparing the efficiency of private health service providers and increasing competition between them. Some may gain but others may lose. Bearing in mind this risk, health service providers may refuse to share their data. Mandatory data sharing would be the only way to overcome private disincentives to share data for this purpose and realize the economies of scope in data aggregation.

(b) Incomplete contracts, risks, transaction costs and missing data markets

Dosis and Sand-Zantman (2019) distinguish between contractible and non-contractible data rights. **Incomplete contract theory** says that *ex-ante* contracts are necessarily of

finite length and can never include provisions for all possible *ex-post* events or unforeseeable events.[30] Some contractual provisions may also 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 items like risks from data leaks, or unexpected costs and benefits not foreseen in the contract. In traditional contracts, unexpected costs and benefits are assigned to the owner of the traded good or service. Since there are no legally defined ownership-type of rights for data, this makes it difficult to apply that solution. Moreover, bilateral contracts cannot be enforced vis à vis third parties. If data leak to a third-party, the original data collectors become exposed again to non-excludability risks. This may weaken incentives for data sharing.

One solution to incomplete contracts is to appoint a third party to settle the case in such an event, usually a judge. This may involve costly court proceedings. A more efficient solution may be to **assign ownership-type of rights**. 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. There are no *in rem* ownership rights in data (Duch-Brown et al., 2017). Debates on the possible introduction of such rights (European Commission, 2017a) have faltered and attention has now shifted to introducing data access rights (European Commission, 2018a, Drexl, 2018). Ownership and access rights are complements. The first is a residual exclusion right, the second a specific inclusion right. Yet, the question remains who should get such rights, if any.

In the absence of legal ownership rights, **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 demonstrated the risks of bilateral contracting for non-rival data[31].

A related dimension is transaction costs in data exchanges. Negotiating, writing and monitoring the implementation of a contract is costly. Writing a contract that foresees all possible cases would be infinitely costly. Contracting parties want to limit their negotiation and implementation cost. Contracts therefore come inevitably with residual uncertainties that can give rise to more costs during monitoring and execution of the contract (Williamson, 1985). **High transaction costs can especially occur in non-rival data markets where *ex-ante* transaction costs and *ex-post* implementation risks can be very high**. They can lead to market failure in access to data. The higher

[30] Also Deloitte et al. (2018).

[31] Facebook was lax in enforcing the contractual agreement with Cambridge Analytica that the data should have been deleted, not re-used beyond the agreed timeframe and for other purposes. Lax enforcement backfired on Facebook.

the cost of contracting in the market, the more firms will want to circumvent the market and keep exchanges in-house (Coase, 1937). Building a protective data wall around the firm becomes a private response to data market failures. However, too much data accumulation inside big firms poses competition problems and a new source of market failure.

This can be illustrated with an example from personal data markets. Acemoglu et al (2019) argue that the market value of an individual's personal data is low and depressed because of positive externalities in data aggregation across many individuals. The aggregator has a large dataset that has predictive value for the profile and preferences of individuals who are not in the dataset. Individuals can protect and refuse to share their private data but the aggregator can still target them with advertising because his overall profile data are sufficiently accurate, based on predictive data obtained from similar individuals who were willing to share their data. The very low and depressed value of personal data also explains why natural persons have little incentive to invest in the management of their personal data, despite the availability of a variety of tools for the purpose. **Transaction costs are higher than the potential benefits from personal data management**. This results in a missing, or at least very much reduced, market for privacy and Personal Data Management Services.

(c) Asymmetric and imperfect information

Imperfect and asymmetric information between individuals and large data collectors are almost natural states in a data-abundant digital world. For example, online platforms as data aggregators will always have more and better information on the services markets that they cover, compared to users of these platforms (consumers and sellers). The EU Platform-to-Business Regulation[32] requires a platform to inform businesses about the extent of access to data that business users or consumers provide for the use of the platform's services or that are generated through the provision of those services. Apart from this form of information provision, the Regulation does not offer businesses a right to access data. Even if businesses would be offered access to data related to their own activities on a platform, this gives them only a subset of the total market information that the platform has. It does not restore the informational level playing field between platforms and users.

Platforms provide users with search engines and advertising channels to guide the matching process between sellers and buyers. Platforms have an interest in increasing the efficiency of the matching process because it draws more users to the platforms and increases their own revenue. At the same time, they seek to maximize their own profit and will distort the available information to achieve that goal. They draw an information wedge between the interests of users on both sides of the market to advance their own interest as platform operator (De los Santos and Koulayev, 2017). As a result, platforms do not necessarily use the available data to optimize the combined social welfare of all users and the platform operator. The stronger their market position, the more they may distort the information picture.

[32] Regulation 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services [2019] OJ L 186/57, art. 9.

3. Third-party intermediaries to reduce market failures

In the previous section we identified and discussed several sources of market failures in B2B data markets where the profit-maximising behaviour of private data holders prevents maximising overall social welfare for society – e.g. the combined welfare of data holders, potential re-users who could benefit from access to the data, and consumers who would benefit from improved services. The next question is: what can be done to overcome these market failures? Third-parties may act as **intermediaries and apply new technologies and ways of organizing markets that will reduce and remedy previous market failures** and enable transactions that were previously not feasible.

They can be private for-profit firms or non-governmental not-for-profit organisations. They can be community-based organisations with a relatively stable group of users who participate in decision making, or private entities with variable users who do not participate in decision making but is attracted by the services offered. An extensive "commons" literature covers the community type (Madison et al., 2016; Bertacchini et al, 2009). They can also be called **"technical enablers" in data sharing** (European Commission, 2018b). They can be neutral with respect to data uses or they can be an active stakeholder in the value-added generated by the data. De Streel and Tombal (2020) present a typology of third-party intermediary platforms along two dimensions: the degree of active intervention of the third-party and the degree of multi-laterality among the participants.

Governments could intervene to create more favourable conditions for market-based facilitation. The European Commission's Data Strategy (2020) suggests to put in place an enabling legislative framework for the governance of common European data spaces in order to facilitate data use, to prioritise standardization activities and to foster data interoperability. The Commission also intends to invest in establishing EU-wide common, interoperable data spaces with standardisation, appropriate data sharing tools and governance frameworks to promote interoperability and data sharing in strategic sectors and domains of public interest. Creating common governance institutions implies the introduction of third-parties into the data market, legal entities that act as intermediaries. In this section we discuss the possibilities and limitations of these intermediaries.

From an economic perspective, third-parties can facilitate B2B data sharing between firms in several ways. They can (i) reduce *ex-ante* transaction costs and *ex-post* contractual risks transaction costs, (ii) overcome coordination problems, (iii) act as commitment enforcement devices and facilitate self-regulation between private agents, and (iv) set interoperability standards. We can also apply the two interpretations of economies of scope that we developed in the previous section to classify third-party enablers in two groups:

- those who facilitate vertical interaction between upstream data holders and downstream re-users for the production of data-driven services, essentially (i) and (iv),
- and those who facilitate horizontal coordination among data holders for the purpose of realizing economies of scope in data aggregation, mostly (ii) and (iii).

The split between these categories may be a bit artificial but is useful to illustrate their roles. In reality, there may be hybrid versions of all these typologies.

3.1. Reducing risk

As already explained, once data holders and re-users have agreed on a contract that defines the conditions for access and re-use of data, implementation of the **contract may entail significant risks**. For example, upstream mobile phone operators may agree to share data with downstream transport and mobility service providers for the purpose of improving traffic and mobility management in a city. The data provider will not want to share primary phone user data but only derived mobility indicators at a sufficiently aggregate level that protects the privacy of its clients. This does not require a third-party intermediary. The simple solution is indirect data sharing whereby the mobile phone operator produces the derived dataset that is delivered as a value-added service to mobility service provider. However, if the mobility service provider requires pooled data from all mobile phone operators in a city in order to create more detailed insights than a third-party intermediary may be necessary. None of the data suppliers wants the others to access their primary customer data. None of the participants will trust the other with collecting the data because each has an interest in getting the full dataset. Solving this problem **requires a trusted third-party intermediary** who collects the data on his server platform, performs the analysis and ensures that only the processed results are shared with users.

Data trusts[33] and industrial data platforms fit this profile. 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. In this way, neutral third-party intermediaries overcome data market failures and enable data sharing transactions that would otherwise not materialise because of perceived risks[34].

3.2. Transaction costs and standardisation

Post-contractual risk-reduction assumes that contracts have already been agreed between data holders and re-users. However, prohibitively **high transaction costs may occur prior to the signature of a data sharing contract** because of search costs for appropriate data and partners and the cost of negotiating a contract between the parties. Overcoming pre-contractual transaction costs requires a more active intermediary who has a stake in reaching B2B data deals between providers and re-users but has no stake

[33] See also the Open Data Institute on data trusts https://theodi.org/project/data-trusts/

[34] There are many historical examples where third-party intermediaries fulfilled trustee roles in the absence of government-supplied commitment and enforcement services and contributed to economic efficiency and growth (for example, Milgrom, North qnd Weingast, 1990).

in the contents of the data transfer, in order to avoid strategic behaviour by the intermediary. **Two-sided data market places and single-sided data retailers** match data supply and demand and fit this profile.

Reducing transaction costs in data markets is not an easy task. One reason can be that the intermediary management costs are too high compared to the value of the outcome. An example from personal data markets illustrates the problem. The European Data Protection Supervisor has advocated the use of **Personal Information Management Systems (PIMS)**[35]. Despite many start-ups in this domain, there are **no real break-throughs** that have scaled-up to become a significant market player in personal data markets. This can be attributed to the fact that the value of personal datasets from individual consumers is very low because preferences and behaviour of a specific consumer can be inferred from data collected and extrapolated from other consumers (Acemoglu et al, 2019). That reduces the marginal opportunity cost of personal data and pushes it below the opportunity cost of time and effort involved in managing one's personal data in PIMS. Poor cost-benefit parameters erode incentives for individuals to invest in PIMS (Kramer, Senellart and de Streel, 2020:66-73).

Third-party B2B data sharing platforms may be more successful in closed groups with known participants than in open-ended groups of users. For example, the Hitachi B2B data platform in Copenhagen was not successful because data suppliers were reluctant to hand over control over data re-uses to a neutral intermediary platform. Re-use may have negative externality effects on the data supplier, or they may miss out on opportunities to earn more on a data sharing deal. The participants are better off when they have negotiated data sharing conditions with each other; they know what they signed up to and who with.

Despite all these obstacles to reducing transaction costs, there are positive contributions that third-party intermediaries can make to facilitating B2B data sharing deals. Active intermediaries may lower pre-contractual transaction costs by proposing **standardised clauses in data sharing contracts in order to facilitate negotiations**. This seems to have been the intention of the European Commission in its Staff working document (2018b) that proposed non-binding guiding principles for B2B data sharing contracts between data holders and downstream users. These include: (i) transparency about the entities that can access the data, the type and level of detail, and the purposes for which data can be used; and (ii) respect for each other's commercial interests and value contribution to the data[36]. These principles are purely normative guidelines for data traders that have no legal value but could become enshrined in data contracts. Default clauses and principles reduce the costs and difficulty of negotiating a contract. Standard clauses fill a gap in legal provisions regarding data contracts and exchanges.

A question is whether binding legal EU rules are necessary or whether existing general contract and commercial law at the national level provide a sufficient basis to address

[35] https://edps.europa.eu/data-protection/our-work/subjects/personal-information-management-system_en See also European Commission (2020, p 18).

[36] Other principles suggested in the SWD, including ensuring undistorted competition enabling portability to minimise data lock-in, belong more in the sphere of public policy and do not necessarily align with the interests of private actors.

uncertainties with regard to B2B data-based transactions. In fact, the problem may be the lack of precedent rather than a lack of relevant rules. With regard to the relationship between platforms and business users, it is worth referring to the EU Platform-to-Business Regulation which requires providers of online intermediation services to be transparent in their terms and conditions about the access that business users will have to relevant data. [37] Even though the Regulation does not prescribe any contractual duties or rights, higher levels of transparency can be a starting point to facilitate negotiations as well as a basis for determining whether more far-reaching interventions are necessary to stimulate data sharing in platform-to-business relations.

Intermediary platforms may also reduce transaction costs because they can contribute to **standardisation and interoperability of the datasets** that are being exchanged.[38] Standards may emerge (i) bottom-up, driven by market forces that create *de facto* standards, (ii) they may be imposed top-down by regulatory authorities, or (iii) they may be proposed by entrepreneurial intermediaries that can range from single firms to recognised international Standard Setting Organisations (SSOs) that facilitate the emergence of standards in a semi-official way. SSOs help to overcome market coordination failures (Baron et al, 2019).

Voluntary standard setting works when the participants have an incentive to contribute to a standard. For example, in Standard Essential Patents (SEPs) participants in the pool gain because their patents are complements. Pooling makes their individual patents more valuable, compared to their separate use. Standards may fail to emerge when stakeholders have conflicting interests, when their products are competing substitutes, or when some key stakeholders are excluded from the negotiations. For example, a number of car manufacturers proposed the Extended Vehicle Standard for data exchanges between connected cars and aftermarket service users. The latter where excluded from discussions in the standard setting group. The resulting design was perceived as being in the interest of the manufacturers as data holders but undermining the interests of the data re-users, for instance because they exposed commercially confidential re-user information to data holders. As a result, the re-users refused to participate in the coordination exercise (European Commission, 2016).

The **absence of standards may also create entrepreneurial opportunities for firms that offer private solutions** to overcome data market fragmentation and interoperability problems. In the automotive sector for example, intermediary data market places like Otonomo and Caruso ensure data standardisation and interoperability across a large number of manufacturers and service providers. Building appropriate data interfaces entails fixed costs that can only be amortised across a large number of transactions. This implies that services will not be sold at marginal cost but at some cost mark-up.

This mark-up creates some wriggle-room for **third-party intermediaries to become more active players in the data market and seek to capture a larger share of the benefits that they generate as data trade facilitators and transaction cost reducers**. They have an interest in expanding the market by attracting new data

[37] Regulation 2019/1150, art.9.
[38] On data standardization, see generally Gal and Rubinfeld (2019).

suppliers and users and consolidating his market power to charge a higher profit margin. They can shift from being a one-sided data retailer to becoming a two-sided market operator with significant direct and indirect network effects. Strong intermediary market players may impose market-driven standards. This can be illustrated with another example from the automotive sector: the rapidly increasing use of Apple iOS and Google Android operating systems as driver media interfaces in cars sets a *de facto* (dual) standard for car data formats, as defined by these operating systems. Strong consumer demand for these familiar operating systems, also because they facilitate interoperability with other consumer devices, puts pressure on car manufacturers to allow them into their cars. The more widely interoperable data standards that come along with these operating systems may push manufacturers' own standards out of the market.

3.3. Internalisation and redistribution of externalities

In sections 3.1 and 3.2 we limited the role of third-party intermediaries to reducing transaction costs and risks to facilitate data re-use in a vertical relationship between data collectors and re-users. In this section we discuss how third-party intermediaries can capture externalities that occur in the wake of economies of scope in data aggregation in a horizontal relationship among data collectors.

Data holders and re-users may not be able to reach an agreement when they fail to capture (and monetise) the gains from data sharing or do not agree on the distribution of the gains. These are typical anti-commons coordination problems: market segmentation fails to capture the positive externalities of economies of scope in data aggregation. Segmentation may occur because parts of the dataset are controlled by different parties, or because data holders and data sources each have their own claims over the data. **Third-party intermediaries** may help to overcome these hurdles by offering a business model that can **capture and monetise the externalities, and by setting up a mechanism for redistribution of the gains** to bring on board the parties that feel unfairly treated or are at risk of losing out in data sharing.

Search engines are a good example of the role that a third-party intermediary can play in capturing the gains from data sharing. We already pointed out that the value of personal datasets from individual consumers can be very low because preferences and behaviour of a specific consumer can be inferred from data collected and extrapolated from others. Consumers are willing to trade them in for search services that are more valuable for consumers than their personal data. Search engines subsidise this free search service with an advertising business model that captures and monetises a substantial part of the externalities that come with the aggregations of low value personal data across many consumers (Brynjolfsson, et al, 2019; Acemoglu et al, 2019). The intermediaries are no longer neutral players in this setting. They are actively involved in extracting and redistributing the gains from data sharing and have a stake in the process.

More generally, **intermediaries that process data and extract value-added from this processing usually become non-neutral and more active third-parties**. For example, platform intermediaries can collect relatively low value data inputs from many users on one side of the market and aggregate them into large sets from which they can extract valuable insights through machine learning, and sell these insights to customers on another side of the market. Data input providers will naturally be in a weak bargaining

position because they cannot realize the economies of scope from data aggregation on their own. This may give rise to perceived imbalances in data market power (European Commission, 2020, p 12).

The **benefits of data pooling are often very difficult to monetise because they constitute non-excludable public goods**. For example, pooling health data from across health service providers may contribute to identifying strategies that can improve public health. This positive externality is difficult to monetise for an individual health service provider who therefore has no financial incentive to participate in a health data pool. This may lead to different responses. In Finland for example, the **government intervened to overcome this market failure by making health data sharing mandatory** for certain private and public health service providers and pooling the data in a government agency that makes them accessible to medical researchers who have no obligation to share any externality benefits with the data contributors. [39]

Alternatively, a **third-party may set up a business model that enables the monetisation of these benefits and channel at least some of the benefits back to the data contributors**. For example, in the UK, Google Health Services (DeepMind) reached an agreement with a government health services group to access a vast pool of personal health data. In return, the medical service provider and patients will benefit from improved and free health advisory services provided by Google DeepMind. At the same time Google can monetise these non-rival insights by providing paid health advisory services to parties that do not participate in the programme (Powles and Hobson, 2017). Google Health is not a neutral intermediary between patients and health service providers. It has an incentive to monetize these new health services beyond the participants in the data pool. The advantage of the Finnish option is that it makes health data available to a much larger group of firms and researchers and not a single firm.

Inequality in the pay-off matrix between participants may become an obstacle. For example, sharing mobility data between transport service providers in a city may have a positive expansion effect but also a negative substitution effect on the revenue and market share of data contributors (Carballa-Smichowski, 2018). The net effect may be uncertain at the time of decision. In order to bring hesitating participants on board in the data pool, an **internal compensation and re-distribution mechanism may have to be set up**. A growing pool of benefits can more easily be re-distributed to compensate potential losers: all participants still gain[40]. Market failure in data aggregation may arise from equity concerns in this case.

3.4. The role of third-party intermediaries

We argue in Section 2 that it is unlikely that either exclusively private data use or fully shared common data pools are optimal data governance regimes. In most situations, an

[39] More details regarding the FINDATA Act can be found here https://stm.fi/en/secondary-use-of-health-and-social-data

[40] In economics jargon: a technological change may not be strictly Pareto welfare-improving when some participants gain while others loose. It may however be Kaldor-Hicks optimal when the winners could hypothetically compensate the losers and still be better off. The emphasis is on "hypothetical".

optimal solution is likely to be somewhere in between. B2B data sharing markets already cater to some extent to data sharing demand but they may fail to produce a socially optimum amount of sharing, for a variety of reasons explained in Section 2. In this section we explored what third-party intermediaries can do to overcome some of these market failures. We concluded that there may be some **low-hanging fruit that these intermediaries can easily collect, such as standardised data formats and contracts, acting as a commitment device for the implementation of contracts and reducing ex-post risks**. They have a harder time overcoming ex-ante transaction costs, unless there are economies of scale in these costs. Third-parties face many hurdles related to finding data sharing partners and the negotiation and management of complex contracts.

This is the domain of **semi-commons or governance agreements** that seek to overcome the pitfalls of commons – that lead to over-utilisation and under-investment and facilitate free-riding – and anti-commons – exclusive private use that leads to underutilization and keeps data locked in silos. Smith (2005) argues that semi-commons are often costly solutions to overcome the pitfalls of both extreme regimes. Whether they are worth the cost depends on the value of the agreement. We give examples where they are too costly and fail to overcome the market failure that they try to address.

Smith (2005) discusses the example of US telecommunications infrastructure semi-commons whereby incumbent telecommunications operators are forced to share their non-rival infrastructure with new start-ups at a fixed price. This leads to strategic behaviour by incumbents and start-ups and results in underinvestment in infrastructure. Getting the price right is a key variable in striking an appropriate balance between parties in this semi-commons governance setting. Price setting is handled by a complex and costly institutional set-up involving a supervisory authority (third party intermediary), courts (third-party dispute settlement), etc. By contrast, giving exclusive private property rights is much cheaper to manage – the exclusive owner sets the price – and so is the commons – there is no price. Finding more cost-efficient intermediary governance systems would open up new B2B data market opportunities, enable the emergence of so-far missing markets.

In the next section we discuss government intervention in B2B data markets and distinguish between *ex post* case-based competition interventions and *ex ante* general regulatory interventions. Governments can set up governance structures, as in the case of telecoms for example, or patents and copyright regimes.

4. Competition Law and Regulation to remedy market failures

In Section 3, we outline that third-party intermediaries may contribute to solve market failures, by reducing transaction cost and risks, overcoming coordination problems and acting as commitment devices between private agents to facilitate the setting of interoperability standards, possibly through co-regulation. Yet, that may not be sufficient to solve all data market failures and more active public intervention may be necessary. This could first be done with competition policy. In this case, competition authorities intervene *ex post* to correct market failures on a case-by-case basis (Section 4.1). However, these market failures could occur on a wider scale and on a regular basis, and competition law might not be sufficient. In such cases, it might be necessary for legislators to set *ex ante* mandatory rules that reduce the *de facto* exclusive control of data holders and allocate rights to other parties. Legislators may impose obligations on data holders and assign legal rights to stakeholders in the data market, ranging from full and exclusive ownership rights to data, to more specific and limited access or portability rights (Section 4.2).

4.1. *Ex post* Competition enforcement

If the market failures mentioned above occur only occasionally and can be easily solved, competition authorities may be better placed to address these issues on a case-by-case basis. As regards the overall goal, the Court of Justice clarifies that the EU competition rules are necessary for the functioning of the internal market and seek 'to prevent competition from being distorted to the detriment of the public interest, individual undertakings and consumers, thereby ensuring the well-being of the European Union'.[41] To this end, **competition law protects consumer welfare and to contribute to internal market integration** by intervening against restrictive practices, abusive behaviour and concentrations that significantly impede effective competition.[42]

As regards competition interventions to address data market failures, the two most relevant scenarios are interventions in the context of abuse of dominance to mandate access to data where a dominant firm refuses to do so and interventions to address discriminatory prices or otherwise unfair conditions charged for access to data by a dominant firm.

With respect to **refusals to give access** to data, the so-called **essential facilities doctrine** is relevant. Refusals to deal are a type of abuse of dominance under Article 102 TFEU. The consequence of a finding of such an abuse is that a dominant firm is obliged to enter into dealings with third parties. As such, the essential facilities doctrine mandates dominant firms in certain circumstances to provide access to inputs or assets that are essential for competitors in order to enter a related market. The doctrine has been developed in a long line of cases dealing with access to physical infrastructures as well as

[41] See for instance Case C-52/09 *TeliaSonera Sverige*, ECLI:EU:C:2011:83, par. 22.

[42] We will leave the two branches of EU competition law targeting state aid and public undertakings out of our consideration.

licensing of intellectual property rights.[43] The application of the doctrine to data has been a topic of discussion in literature (Colangelo and Maggiolino, 2018; Drexl, 2017; Feasey and de Streel, 2020; Graef, 2016) as well as in competition policy reports (Crémer et al, 2019; Schweitzer et al, 2018). Despite the attention for the issue, no competition interventions have occurred so far at EU level to force access to data in order to open up data markets in the sense we focus on here. At the national level, competition interventions have taken place to force dominant firms to give access to datasets in order to enter more traditional markets, including those for the supply of gas[44] and lottery services[45]. Although the conditions of the essential facilities doctrine (namely indispensability, exclusion of effective competition, new product and absence of objective justification) can generally be applied to data, some limitations are worth mentioning.

First, the non-rivalry and wide availability of data will often render a dataset non-indispensable as the Commission has concluded in a number of merger decisions.[46] Second, an important restriction in the way the essential facilities doctrine is currently interpreted in case law is that there must be 'leveraging' by the dominant firm. This involves the use of the firm's dominant position in the market for the input in order to gain a stronger position in the related market in which the access seeker wants to introduce a new product or service. In previous cases, the EU courts have indeed interpreted the condition of exclusion of effective competition in a way that the dominant firm must itself already be active in the related market.[47] This means that scenarios in which access seekers want to use the input for a product or service which is situated in a different antitrust market than the one where that the dominant firm is active fall outside the scope of the essential facilities doctrine as currently interpreted (Drexl, 2017; Graef, 2016). Yet, these are the precise scenarios that are to be expected in data markets because of the many ways in which data can be used for new purposes. In a growing data economy, market players will likely come up with new uses of data going much beyond the products and services offered by the holders of data.

Absent a more expansive interpretation of the essential facilities doctrine, dominant firms will thus be able to prevent such new uses of existing datasets by refusing to share data with third parties. More generally, the distinct characteristics of data and the competitive dynamics of the data economy may require a different balance to be struck between the

---

[43] See among others *Telefis Eireann and Independent Television Publications Ltd v. Commission of the European Communities (Magill)*, joined cases C-241/91 and C-242/91, [1995] ECLI:EU:C:1995:98; *Oscar Bronner GmbH & Co. KG v. Mediaprint Zeitungs*, case C-7/97, [1998] ECLI:EU:C:1998:569; *IMS Health GmbH & Co. OHG v. NDC Health GmbH & Co. KG*, case C-418/01, [2004] ECLI:EU:C:2004:257; *Microsoft*, case T-201/04, [2007] ECLI:EU:T:2007:289.

[44] French Autorité de la concurrence, Decision n°17-D-06 (*GDF Suez*), 21 March 2017, available at http://www.autoritedelaconcurrence.fr/pdf/avis/17d06.pdf.

[45] Decision 2015-P/K-27 of 22 September 2015 of the Belgian Competition Authority, *Stanleybet Belgium/Stanley International Betting and Sagevas/World Football Association/Samenwerkende Nevenmaatschappij Belgische PMU v. Nationale Loterij*.

[46] See among others Case COMP/M.4731 – *Google/ DoubleClick*, 11 March 2008; Case COMP/M.7217 – *Facebook/WhatsApp*, 3 October 2014; Case M.8788 – *Apple/Shazam*, 6 September 2018.

[47] See *Tiercé Ladbroke v. Commission*, case T-504/93, ECLI:EU:T:1997:84, par. 133. The case dealt with a refusal by organisers of French horse races to provide Ladbroke, who was offering betting services in Belgium, with a transmission license for sound and pictures of the French horse races. Apart from the lack of indispensability, the General Court held that the condition of exclusion of effective competition was not met because the organisers of the French horse races were not competitors of Ladbroke on the relevant market for the provision of betting services in Belgium and could therefore not be seeking to reserve that related market for themselves.

costs and benefits of a competition intervention, possibly calling for a lower threshold to be applied to mandating access to data as compared to access to other products.[48]

An expansion of the interpretation of the essential facilities doctrine beyond cases of leveraging and a lower threshold for the condition of indispensability (as well as new product)[49] more generally seems feasible, as competition law is flexible and can adapt its application to the relevant circumstances of the market. The essential facilities doctrine itself has developed over a number of cases in which nuances in the application of the conditions were made. Once a precedent is set in which a competition authority or court imposes a duty to give access to a dataset based on the essential facilities doctrine, this will clarify the extent to which *ex post* competition enforcement is capable of addressing data market failures originating from dominant firms refusing to share data with other market players. Even though competition enforcement takes place *ex post*, it can certainly also have an impact *ex ante* on the future strategy of market players in providing access to data they hold.

Apart from refusing to supply data, market failures may originate in the **conditions under which a dominant firm decides to give access to data**. As evidenced by the Amazon investigation that the European Commission opened in July 2019,[50] a vertically integrated firm may discriminate against competitors by providing its own downstream business with preferential access to data. While there are efficiency reasons for a vertically integrated firm to keep the use of data on the downstream market to itself, this may not be the optimal outcome from a welfare perspective. A competition intervention can then address the anticompetitive effects resulting from such forms of self-preferencing.

Finally, specific disputes could pertain to one of the **conditions under which a dominant firm decides to give access to data, namely the price** of the access.[51] In this regard, Drexl (2017) points out that the *Huawei*[52] judgment, where the Court of Justice of the EU created a negotiation framework for the licensing of Standard Essential Patens on FRAND (Fair Reasonable And Non-Discriminatory) terms[53], could be used as inspiration for cases of data access. This negotiation framework, applied to data, could be the following as explained by Tombal (2020): Once the access seeker has expressed its willingness to pay a fair remuneration for the data, the data holder must present a specific written offer specifying the price and the way in which it is to be calculated; It is then for the access seeker to respond diligently to that offer in good faith and without delaying tactics; Should the access seeker not accept the offer made to it, it has to submit to the data holder, promptly and in writing, a specific counter-offer that corresponds to a fair remuneration; and Where no agreement is reached on the details of

[48] See for instance Schweitzer et al (2018) who argue that: 'There are good reasons to think that, depending on the exact setting, the threshold for finding that a refusal to supply data constitutes an abuse may be somewhat lower than the threshold for finding an abuse in cases of a refusal to grant access to infrastructures or to intellectual property rights. This is true in particular if and to the extent that the refusal to grant access relates to data which is generated virtually incidentally and without special investment'.

[49] For a more extensive analysis, see Graef, Tombal and de Streel (2019).

[50] Press release European Commission, 'Antitrust: Commission opens investigation into possible anti-competitive conduct of Amazon', 17 July 2019, available at https://ec.europa.eu/commission/presscorner/detail/en/IP_19_4291.

[51] On the value and the price of data, see also OECD (2013); Feasey and de Streel (2020).

[52] Case C-170/13, *Huawei v. ZTE*, ECLI:EU:C:2015:477.

[53] Case C-170/13, *Huawei,* paras. 60-69.

the remuneration following the counter-offer, the parties should, by common agreement, request that the price be determined by an independent third party. To set this price, this independent third party could, for instance, rely on the "baseball arbitration" mechanism, originally used in the USA for baseball salary negotiations.[54] According to this mechanism, each party proposes a price to the third party, who is tasked with choosing the price that appears to be the most "reasonable" under the circumstances of the cases. This forces each party to be restrain themselves from proposing unreasonable prices, as they know that if they suggest an extravagant price, while the other party suggests a more "reasonable" one, the independent third party will pick the other party's price. A variant form is the "night baseball arbitration", where the independent third party first decides itself what could be a reasonable price and then looks at the party's proposals and choses whichever is the closest to the price it first considered.

Tombal (2020) notes that this independent third party could be the Support Centre for data sharing, created in 2019. This Centre could be in a good position to assess the appropriateness of the conditions of a data sharing agreement since it is tasked with collecting the best practices and existing model contract terms.[55] Thus, it should have a basis of comparison to assess the prices proposed by both parties. However, as the Support Centre for data sharing has been conceived as a recommendation/information service, procured by the Commission, on the existing data sharing practices, it would likely lack the legal power to impose a price. To do so, an existing regulatory may be to have this task or a new body may need to be created, but could nevertheless feed on the knowledge accumulated by the Support Centre for data sharing.

The current lack of competition interventions to mandate access to data, and the time it will take to clarify the conditions under which refusals to give access to data are abusive, may justify looking for remedies beyond *ex post* competition enforcement. The Commission is considering a **New Competition Tool** allowing competition authorities to intervene before a competition law infringement has been established, including for mandating access to data. This is a **hybrid form of intervention**, which lies in between ex post competition enforcement and ex ante regulation. However, the future availability of such a hybrid intervention in the toolbox of competition authorities does not solve the issue of having to establish a threshold for intervention, just as would have to be done under the essential facilities doctrine.

4.2. *Ex ante* regulation

While reliance on competition law could prove useful to tackle some of the market failures identified in Section 2, this regime has its limits. Indeed, competition interventions only take place *ex post* on a case-by-case basis. Litigation may be costly and can take a long time to reach a final outcome in courts. Surely, once this outcome has been reached, this will inevitably also have an impact *ex ante* on how market players operate. By acting against new types of behaviour for which the application of the rules is not yet clear, the boundaries set by the competition authority will also apply to other firms, even in other sectors. Such a competition intervention thus sends a signal to the market about what behaviour will be accepted. However, this does not prevent the undertakings from knowingly acting in a way that infringes competition law and if they do

[54] See http://www.arbitration.com/articles/what-is-baseball-arbitration.aspx.
[55] https://eudatasharing.eu/

so, the harm may often be done before the case can be resolved. This is especially true in "winner takes all" markets, where it becomes extremely difficult to contest the incumbent's position once the market has tipped in its favour. To avoid this, regulators can intervene *ex ante* to anticipate problems. Moreover, regulatory intervention is also justified if these market failures occur on a wider scale and on a regular basis and when the design and the monitoring is very complex and require an effective governance framework that competition law generally does not offer. As explained in Feasey and de Streel (2020), this warrants a general or sectoral regulatory policy answer.

Proposals have been made to require market players to **share data as an *ex ante* requirement to prevent market tipping, ensure market contestability and stimulate innovation** (Kramer, Schnurr and Broughton-Micova, 2020; Prüfer and Schottmüller, 2017), which are objectives going beyond those underlying competition law as such. A suggested modality for such a data sharing framework is to require market players to share an amount of data in line with their market share (Mayer-Schönberger & Ramge, 2018; Prüfer, 2020).[56]

As explained in the EU Support Centre for Data Sharing (2020), the **EU legal framework already contains several rules imposing or encouraging the portability and the access of personal and non-personal data**. Some rules are horizontal and are mainly composed of: for personal data, the General Data Protection Regulation (GDPR)[57]; for non-personal data, the Digital Content Directive (DCD)[58] applicable in a B2C relationship and the Free Flow of Data Regulation (FFDR)[59] applicable in a B2B relationship. Other rules are sectoral and impose data access in particular in: the financial sector with Second Payment Services Directive (PSD2) imposing access to payment account data, which has been completed in the UK with the Open Banking Programme;[60] the automotive sector with the new Motor Vehicle Regulation imposing access to some vehicle data;[61] the energy sector with the new Electricity Directive imposing access to some customers data.[62]

[56] Mayer-Schönberger and Ramger (2018:167): 'we suggest what we term a progressive data-sharing mandate. It would kick in once a company's market share reaches an initial threshold – say, 10 percent. It would then have to share a randomly chosen portion of its feedback data with every other player in the market that requests it. How much data it must make available would depend on the market share captured by the company. The closer a company is to domination, the more data it would have to share with its competitors'.

[57] Regulation 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, and repealing Directive 95/46 (General Data Protection Regulation), OJ [2016] L 199/1.

[58] Directive 2019/770 of the European Parliament and of the Council of 20 May 2019 on certain aspects concerning contracts for the supply of digital content and digital services, OJ [2019] L 136/1.

[59] Regulation 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union, OJ [2018] L 303/59.

[60] Directive 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, OJ [2015] L 337/35, arts.66-67.

[61] Regulation 2018/858 of the European Parliament and of the Council of 30 May 2018 on the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles, OJ [2018] L 151/1, arts.61-66.

[62] Directive 2019/944 of the European Parliament and of the Council of 5 June 2019 on common rules for the internal market for electricity [2019] OJ L 158/125, art.23.

Table 1: EU legal framework for data portability and access

	Personal data	Non-personal data
Horizontal	- GDPR (2016)	- DCD (2019) in B2C - FFDR (2018) in B2B
Sector-Specific	- Financial: PSD2 (2015) - Automotive: Motor Vehicle Regulation (2018) - Energy: Electricity Directive (2019)	

(a) Data portability rights for the data subject

(i) EU regulatory framework on data portability

Portability rights were first established for personal data under the EU GDPR[63]. It creates an unequivocal link between the data controller and data processor and the data subject, grounded in data protection as an inalienable human right. The data subject holds partial but inalienable rights to these data - no transferable ownership rights - including the right to access and delete their personal data, and to port the data under the relevant conditions specified in the GDPR. Data holders face resale and reuse restrictions that can only be lifted if they have a lawful basis to do so.[64] Indeed, according to the *purpose limitation principle*, personal data can only be processed for specified, explicit and legitimate purposes, and cannot be further processed in a manner that is incompatible with those purposes.[65] This means that data that has been collected for a specific purpose cannot be shared with third parties if this act of sharing does not fit within this initial purpose. Because the resale or reuse is a new processing activity, distinct from the initial one, it requires a new lawful basis.[66]

The **data subject can use her portability rights under Article 20 of the GDPR to take the initiative to facilitate reuse**. At the request of the data subject, a data controller is obliged to transfer the data to the data subject, who can then upload the data herself to another data controller, or directly to a third-party controller of the data subject's choice. The latter option only applies in cases where such a direct transfer is 'technically feasible'. This opens the door for a roundabout B2C2B way to achieve B2B data sharing in the case of personal data. An example of this is the Italian start-up

[63] As explained above, the GDPR is based on fundamental human rights considerations; it is not a response to a perceived market failure in personal data markets. It does however have an economic impact on data markets.
[64] Article 6 of the GDPR. Consent is only one of these six lawful bases.
[65] Article 5.1.b) of the GDPR.
[66] This lawful basis can only be consent or a legal obligation (Art. 6.4 of the GDPR).

company Weople that implements data portability requests on behalf of data subjects in exchange for the provision of benefits proportional to the amount and quality of personal data transferred to the platform. Weople then merges this information into its own database to pursue its own commercial purposes of data enrichment.[67]

As pointed out by the European Data Protection Board (called Working Party 29 at the time) in its guidelines on the right to data portability, Article 20 GDPR aims to strengthen the control that the data subjects have on their own personal data, as it represents an opportunity to re-balance the relationship between data subject and data controllers, and this, by affirming individuals' personal rights and control over the personal data concerning them[68]. Moreover, the data portability right also empowers data subjects by making it easier for data subjects to switch between service providers and thereby stimulate competition between service providers (multi-homing in consumer platforms).[69]

A similar debate about the character of the right to data portability is present in literature. One view is that data portability fits with the fundamental rights nature of the data protection regime because it promotes individual control over personal data and thus enhances informational self-determination (Lynskey, 2017). Another view is that data portability can be better characterised as enabling free flow of data among controllers through its use as a tool for individuals to switch where access to data is crucial for competition (Drexl, 2017). Discussions about the character of the right to data portability are not merely of theoretical value but can impact on how its scope is interpreted by data controllers, data protection authorities and courts in practice. This is especially relevant for the limitations in the scope of the right to data portability.

While the GDPR can thus be considered as an appropriate legal instrument to facilitate indirect B2B data sharing through portability rights for personal data, a **number of limitations in the scope of the right to data portability** need to be considered as pointed by Kramer, Senellart and de Streel (2020). First, the scope of this right is limited to specific categories of personal data processing. Indeed, data subjects can only call upon their data portability right for processing carried out by automated means, and which are based either on the data subjects' consent or are necessary for the performance of a contract.[70] There is thus no general right to data portability.

Moreover, the scope of the right to data portability is limited to certain specific categories of personal data because data subjects only have the right to receive the **personal data concerning them, which they have "provided" to a controller**.[71] In this regard, the European Data Protection Board identifies three categories of personal data and

[67] In August 2019, the Italian Data Protection Authority asked the European Data Protection Board to provide advice on the legality of this use of the GDPR's right to data portability: https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9126725.

[68] Working Party 29, *Guidelines on the right to data portability*, WP 242 rev.01, 13 April 2017, p. 4.

[69] In its first version of its guidelines, the Working Party 29 even indicated that this was the primary aim of this new right, as it should facilitate the creation of new services: Working Party 29, *Guidelines on the right to data portability*, WP 242, 13 December 2016, p. 4. However, this indication was deleted in the revised version of April 2017, which now states that that the main objective of this right is to promote data subject empowerment and that the GDPR aims to regulate the processing of personal data, and not to deal with competition law issues.

[70] Article 20(1) of the GDPR.

[71] Article 20(1) of the GDPR.

considers that only the first two should be considered as data "provided" by the data subject.[72]

The first category of personal data covered by the right to data portability are "data actively and knowingly *provided* by the data subject". This includes, but is not limited to, an email address, user name, age or any other information that may be provided, for example, by completing an online registration form for a service, social network, website, etc.

The second category of personal data covered by this right are "*observed data* provided by the data subject by virtue of the use of the service or the device". Examples include the search history of a data subject, the history of the websites he or she has visited, or traffic and location data generated by the use of a mobile application.

The third category of personal data, namely "*inferred data* and derived data created by the data controller on the basis of the data provided by the data subject" will, on the other hand, not be covered by the data portability right. This refers to data resulting from a subsequent analysis carried out by the controller on the basis of raw data provided (actively or observed) by the data subject, such as user profiles created by the controller on the basis of the analysis of data provided by the data subjects, or the results of an assessment of the data subject's health based on the health data collected by his smart watch[73].

Moreover, as pointed out in Article 20.4 of the GDPR, this right to data portability needs to be articulated with the rights and freedoms of others, that it shall not affect. This articulation must be done with two categories of data, namely the personal data of other data subjects, on the one hand, and data protected by trade secrets or intellectual property rights of the data holder, on the other hand.

Finally, a problem with Article 20 GDPR, with regard to its ability to remedy data market failures, is that it is not (yet) sufficiently operational. It defines minimal interoperability requirements but arguably does **not allow for real time access and does not foresee process interoperability and technical compatibility** (Crémer et al, 2019; Kramer et al. 2020). It is therefore difficult to apply in more dynamic B2B data sharing settings where time is a critical factor. This brings us back to the lack of interoperability as a source of market failure, as discussed in Section 2.

A question is how far the scope of the right to data portability can be stretched to make it more effective as a remedy for data market failures, while staying true to its data protection origin. One could argue that existing and more far-reaching sector-specific interventions may influence the interpretation of the GDPR in a way that will require real-time access in sectors where the necessary infrastructure is mandated under sectorial law (Graef et al, 2020). However, as suggested by Krämer, Senellart and de Streel (2020), additional interventions are likely to be more effective for pursuing objectives

[72] Working Party 29, *Guidelines on the right to data portability*, WP 242 rev.01, 13 April 2017, pp. 9-11.

[73] In the interim report published in December 2019 in the context of the market study into online platforms and digital advertising, the UK CMA argued that inferred data can nevertheless constitute personal data and be of value to users from a data protection perspective and to other parties as well including platforms, advertisers and publishers (UK CMA, 2019). Considering the original intention of the EU legislator and the subsequent interpretation by the European Data Protection Board, it seems unfeasible to interpret the right to data portability in a way that will include inferred data. This is also the case because the GDPR stipulates that the right to data portability shall not adversely affect the rights and freedoms of others, which includes the intellectual property of data controllers that may be at stake in inferred data.

relating to competition and innovation with which data protection authorities have less (or even no) enforcement experience and that are not inherently limited to the notion of personal data in data protection law.

Next to the GDPR, the **Digital Content Directive created a data retrieval right for consumers for their non-personal data at the end of a contract**. Article 16(4) of the Directive provides that, in the event of the termination of a contract for the supply of digital content or digital services, the trader shall, at the request of the consumer, make available to the consumer any content other than personal data, which was provided or created by the consumer when using the digital content or digital service supplied by the trader. The consumer is entitled to retrieve that digital content free of charge, without hindrance from the trader, within a reasonable time and in a commonly used and machine-readable format. This resembles, to some extent, the data portability right of the GDPR (Graef, Tombal and de Streel, 2019).

The situation becomes **more complex in the case of non-personal machine data in IoT settings** because there is **no law that establishes an unequivocal legal link between the machine or party that collects the data and the party that would have the right to access the data**. There are no inalienable rights to access machines. The Regulation on the free flow of non-personal data took a first step to overcome these obstacles by suggesting that industry should self-regulate voluntary portability rights for non-personal data. This provision is essentially targeting cloud computing service providers where the data source is usually uniquely identified and subject to a bilateral contract. It seeks to facilitate switching between cloud service providers. There are many other industrial settings however where the identification of rights holders is not straightforward, especially when several parties have supplied data and machines that contribute to a production process.

The European Commission (2017a) suggested the introduction of full data ownership rights for non-personal data as a means to solve the conundrum of who should get access and portability rights. This idea faded when the difficulty of identifying the ownership rights holder was acknowledged (Drexl, 2017; Weibe, 2017; Kerber, 2016; Zech, 2016). The debate has now shifted to assigning access rights instead of ownership rights (European Commission, 2018a, Drexl, 2018). However, the problem of identifying the parties that would receive these rights, and the cases in which such an access should be granted remains the same. In economics, the Coase Theorem suggests that the initial allocation of rights does not matter for economic efficiency when transaction costs are low because initial rights can be traded and end up in the hands of the party that makes the most efficient use of these rights. There are doubts, however, about the applicability of the Coase Theorem to non-rival products like data (Jones and Tonetti, 2019). Social welfare increases when more parties can use data at the same time. Splitting an exclusive ownership right into several non-exclusive access or use rights may thus be more efficient.

(ii) The economic impact of portability

There are already active data trading or data sharing markets, both for personal and non-personal data, even in the absence of data access and portability rights (OECD, 2019). While the data holder seeks to maximize his data revenue, he can ration access to the data to increase their scarcity value and charge a monopoly price that may not be

affordable for the recipient. He will design a market-based data sharing or trading strategy that maximises his private benefits from the data (Bergemann and Bonatti, 2018). As demonstrated in Section 2, that monopolistic strategy does not maximise the volume of data sharing or the social welfare derived from the data unless perfect price discrimination is feasible.

The **introduction of portability rights increases B2B data sharing** between data holders and third-party beneficiaries selected by access rights holders, thereby generates economies of scope in data re-use. Portability puts the data source or originator at the centre of decision-making on data access and re-use. This **works well when the data originator has an incentive** to obtain a data-based service from another party than the initial data collector.[74] For example, in the automotive sector, drivers may want to port their car data to independent service providers because they offer more competitive maintenance services than official dealers. Conversely, it does not work well when these private incentives are missing.

Data portability may strengthen economies of scope in data aggregation. This can occur when a major market player manages to leverage portability rights to collect data from fragmented service providers and aggregate them in a larger data pool that generates stronger economies of scope and efficiencies in service production than the original holders of fragmented datasets. For example, the introduction of Google Android and Apple iOS operating systems in digitally connected cars enables service providers in this ecosystem to lift data out of car manufacturers' brand silos and aggregate them across many car brands and consumer devices that use identical operating systems. The efficiency gains from economies of scope in data aggregation enables the production of new and more efficient services for car drivers that individual car manufacturers cannot match (Martens and Muller-Langer, 2020). Car manufacturers realise this potential and have already initiated cross-brand data sharing to reap the benefits of economies of scope in data aggregation, for example in car navigation services.

At the same time, data aggregation may also lead to increased market concentration in services and efficiency losses because of reduced competition. There are many competing car manufacturers but only a few competitors in car navigation services. In this case, the net social welfare impact of data portability thus depends on the balance between the positive welfare effects from increased economies of scope in data aggregation and the resulting higher service quality and productivity the negative welfare effects from market concentration and decreased competition between alternative service providers. The net effect is an empirical question that may vary across sectors and circumstances.

While portability boosts economies of scope in data re-use, **it may not create a complete level playing field in economies of scope in data aggregation**, because the exercise of the right to portability does not automatically imply an obligation for the controller to erase the ported data.[75] Re-users that want to collect data via portability incur costs to incentivise portability rights holders to initiate a data porting request with the data holder. Some rights holders will accept the incentive offered; others may not. As

[74] See also Kramer, Senellart and de Streel (2020).
[75] Article 20.3 of the GDPR.

a result, the re-user receives only part of the dataset collected by the initial data holder. Consequently, economies of scope in data aggregation are weaker for re-users than for the original data holder. When these economies of scope are important for the service, service efficiency will be lower for re-users. This means that portability cannot completely level the playing field between original data holders and re-users, it is an improvement compared to a situation with no such right. The question is whether this is sufficient to address a possible market failure.

Portability right may also restrict the freedom of the data holder to set a price for the data. Following Article 12(5) of the GDPR, actions taken to comply with data portability requests have to be provided free of charge by data controllers unless requests are manifestly unfounded or excessive, in particular because of their repetitive character in which cases the controller may either charge a reasonable fee taking into account the administrative costs or refuse to act. A question brought up in literature is whether this provision also stands in the way of data controllers charging third parties (not data subjects) for licenses to reuse data in their own services (Graef, Husovec, Purtova, 2019a). That portability is free of charge for a data subject might not imply that reuse of IP-protected data is also automatically free of charge for a third-party controller.

In any case, **portability rights re-allocate data rents from the data holder to the portability right holder and his third-party beneficiaries**. Lower revenue may reduce the incentive for the data holder to collect the data or make the necessary investments in APIs and minimal interoperability. Even when data are a by-product of other operations it may still reduce revenue for the data holder. In a two-sided market model, portability may lower market entry costs for data recipients/users but may increase entry costs on the data supply side. In retail banking for example it might result in higher customer charges for operating a bank account because banks will compensate data revenue losses in other ways.

(b) Mandatory data access right for other parties

The GDPR creates an inalienable link between personal data and the data subject as a natural person who has rights to access his personal data, irrespective of the legal status of the device that was used to collect the personal data. For **non-personal machine-generated data there is no obvious "natural" party that can claim such access rights**. Many parties may have some sort of claim to access the data: machine owners, renters, users, or parties that own the domain where the machine has been working. The machine manufacturer can usually design the machine in such a way that he has exclusive control over access to the data. The legal status of the machine becomes an important issue: Should claims be attached to machine and sensor ownership, to machine users, to data originators or processors, or to third parties? The European Commission (2020) announced its intention to clarify data usage rights for co-generated data in IoT and machine generated data settings.

The example of agricultural data illustrates the difficulties in this regard (Atik and Martens, 2020). A farmer can be the owner or operator of a machine. He can use the machine on his own land or on leased land. Operations with the machine can be driven by data analytics produced by a third-party service provider who has been contracted by

the farmer, the machine owner, the machine manufacturer or the land owner. Each of these parties may claim access to the data. Technical Protection Measures and bilateral contracts will determine who has access to the data - not by right but by technical and economic might. Data access and trading depends on negotiations and market power of the contracting parties. Farmers may feel uncomfortable in this situation and reduce their demand for data-driven machines (Soto and Gomez, 2019). This explains why EU agro-industry organisations proposed a voluntary Code of Conduct for data that seeks to emulate GDPR-like consent, access and portability rights and assign them to farmers. [76] However, the code is ambiguous in the identification of the rights holders. It aims to attribute these rights to farmers as data originators but allows for the possibility to attribute rights to machine owners and operators. Contrary to the GDPR, where data subjects have inalienable rights, the Code makes rights tradable and subject to bilaterally negotiated contracts clauses and market forces.

Portability does not work well when the data subjects have no private incentives to port data to another service provider. In that case, it may be better to shift access rights away from data subjects and assign them directly to data re-users. This requires mandatory access rights for third-parties. Mandatory B2B sharing differs from portability because the data originator has no role to play in the data sharing decisions. Sharing can be initiated directly by the data re-user. This regulatory solution is appropriate when there are clearly defined social welfare gains to be expected while the data holder and originator lack incentives to contribute to these gains. The data holder may actually be in a situation where the costs from sharing may exceed his gains. In that case legislators or regulators may impose mandatory sharing when the net social welfare gains exceed the private costs, although this would entail a welfare re-distribution from data suppliers to recipients.

For example, **in the automotive sector, legislators impose to car manufacturers to share basic maintenance information with maintenance service providers** and ensure open access to the in-car maintenance data socket in order to promote competition between authorised dealers and independent service providers.[77] Manufacturers have no incentive to do so because it weakens the market position of their franchised dealers. Consumer welfare gains from competition in maintenance services justifies this decision. Other examples include sharing of truck parking information between public and private parking operators and transport services providers, and sharing of shipping information between fleet operators and port authorities.[78]

Moreover, portability does not work well when economies of scope in data aggregation are important. As explained in chapter 3 on third-party intermediaries, market players with strong network effects may sometimes be in a position to make attractive service offers to data originators that motivates them to port their data into a pool that generate economies of scope in data aggregation. In the absence of

[76] EU Code of conduct on agricultural data sharing by contractual agreement, April 2018, available at https://copa-cogeca.eu/img/user/files/EU%20CODE/EU_Code_2018_web_version.pdf.

[77] Regulation 2018/858 on the approval and market surveillance of motor vehicles and their trailers.

[78] Commission Delegated Regulation 885/2013 of 15 May 2013 supplementing ITS Directive 2010/40 of the European Parliament and of the Council with regard to the provision of information services for safe and secure parking places for trucks and commercial vehicles. The EU Vessel Traffic Monitoring Directive makes shipping information widely available between firms.

aggregators that can offer these incentives, **legislators or regulators may create a mandatory data pool**.

For example, **B2B health data poolin**g is expected to lead to new insights that can improve and boost the productivity of health services and benefit consumers. Pooling is unlikely to be produced by the market because private health service providers have no incentive to contribute to this pool, unless they can directly benefit from these new insights. Health service providers may incur costs from pooling (data formatting to ensure interoperability, transmission costs) and consumers may fear repercussions for their privacy. In order to overcome this lack of incentives among data sources and holders, the Finnish government made health data pooling mandatory both for public and for certain private health services providers[79], overriding consent requirements for data subjects and commercial confidentiality issues for health service providers. Expected public welfare gains justify this decision. The government created a data pool operator[80] and an oversight board that can grant public and private researchers access to the data. However, in some cases, health data researchers may be in a position to offer benefits to health service providers that create private incentives to contribute to the data pool, as illustrated by the UK health service sharing data with Google Health Services. The Finnish example does not exclude data sharing with private firms that offer direct benefits to the contributors. It avoids, however, that private firms would have to negotiate individually with health service providers because it gives them direct access to the pool of health data. That facilitates competition between private firms that want to research the data.

Another example is **mobility platforms in cities**. Public and private transport service providers have only weak incentives to share their data with competing service providers because it may lead to potentially negative effects on their own business when competitors use the data to set up more efficient competition strategies (Carballa-Smichowski, 2018). Yet, pooling of mobility data may generate economies of scope in traffic management and thereby benefit consumers. Both competition effects and economies of scope in data aggregation play a role in mobility platforms. They require data aggregation across all service providers in order to work efficiently. In order to achieve this, mobility service providers may need to be forced by law to contribute data to a common pool operator in order to overcome disincentives due to possible negative substitution effects between competing transport service providers (Carballa-Smichowski, 2018).

Provisions for data access more targeted at addressing data market failures exist at **sectoral level** to stimulate competition in services, for example in retail banking and payment services, automotive and energy sectors as explained above.

These examples of mandatory data access in order to generate economies of scope in data aggregation raise **questions about the policy decision criterion**. Should it be consumer welfare – the benchmark that is commonly used in competition law –, or should it be social welfare – the benchmark commonly used by public policy economists? Should it be mandatory only in the case of strictly Pareto welfare-improving data sharing where no party loses welfare, or should it also apply to data-sharing that increases

[79] See https://stm.fi/en/secondary-use-of-health-and-social-data
[80] See https://www.findata.fi/en/

overall welfare but may involve welfare losses for some parties (Kaldor-Hicks welfare improving)? For example, can a car navigation service provider direct cars to certain streets in order to find out if the passage is blocked? While this information is very useful for many other drivers, collecting the information imposes a cost on drivers who have been directed to these streets and lost time because of road blocks. Can a navigation app redirect drivers to secondary streets in order to reduce traffic jams in major roads while imposing additional pollution costs on residents along secondary streets? **These are equity and social welfare re-distribution questions that imply a political judgement on the trade-offs between welfare of different groups in society**.

5. Conclusions

In Sections 3 and 4, we examine ways in which market failures in B2B data sharing can be overcome. We first look **market-based solutions through third-party intermediaries** that do not require policy intervention, though "soft" government enabling actions can facilitate the market. Third-party intermediaries can solve market failures when the data sharing parties have an incentive and willingness to move ahead with a deal but face obstacles related to transaction costs, risks and coordination problems.

We then look at situations where the data holder has no incentive to agree to a deal with a data re-user, or may impose monopolistic terms and conditions. In such cases, an intervention by the State to remedy these market failures is required. This could first be done through the means of **competition law**. Here, competition authorities intervene *ex post* to correct market failures on a case-by-case basis.

However, these market failures could occur on a wider scale and on a regular basis, and competition law might not be sufficient. In such cases, **regulators and legislators can intervene by imposing duties to share data** (including in the form of a hybrid intervention through the envisaged New Competition Tool) or by granting mandatory data sharing rights, either to the data source (portability right) or to a re-user (access right). Beyond the existence of market failures as a justification for policy intervention, user empowerment can provide an underlying motivation in particular for improving the effectiveness of portability rights and for the introduction of possible new access rights. Any of such interventions requires the setting of thresholds or conditions that would trigger the application of the obligations. This would involve questions about the type of data involved, the addressees of the new rights/duties and the modalities of access. Existing regulatory instruments and their limitations can provide inspiration for the design of such new interventions. In particular, it is important to map how more proactive implementation and enforcement of available regimes can address current gaps where data market failures remain without an effective remedy.

Portability implies that the data source can take the initiative to request the data holder to transfer the data to the re-user. When legislators make portability mandatory, the data holder cannot refuse the request. The EU GDPR grants a portability right for personal data. However, this right has many limitations as the implementation modalities remain vague and ill-suited to real time transfers and to overcome interoperability problems. There are several examples of more operational sectoral portability rights for personal data in the EU, including in banking, automotive and energy. The GDPR establishes an unequivocal link between a data collector and the data subject as a natural person. For non-personal data such unequivocal links do not exist in EU legislation. As a result, the identification of the data source party is more ambiguous, and may become complicated in IoT settings where several parties collect and contribute data to a data-driven production process.

Portability and access rights can generate economies of scope in data sharing. They may be less effective in generating economies of scope in data aggregation, when pooling of data from several parties is required. Subject to solving coordination problems,

markets can generate these benefits when the parties have an incentive to pool and aggregate data because they benefit directly from the aggregation.

However, when the **participants obtain no direct benefits from the aggregation, mandatory pooling may be necessary to overcome this type of market failure**. Legislator or regulators may also **grant a data access right directly to a re-user, without intervention of the data source**. If such rights are granted for personal data, this requires to comply with the rules of the GDPR, unless the data is anonymised. For instance, the re-user needs a lawful basis to re-use non-anonymised personal data. Indeed, according to the "purpose limitation principle", personal data can only be processed for specified, explicit and legitimate purposes, and cannot be further processed in a manner that is incompatible with those purposes. Moreover, according to the "data minimisation principle", only the adequate, relevant and necessary personal data for the fulfilment of a specific purpose can be processed. This lawful basis could be a legal obligation created by regulators, which would specify the purposes for which the personal data can be accessed and the categories of data covered. For non-personal data, these restrictions do not come into play because there are no legal rights over these data. Nevertheless, there are as yet only few examples of mandatory access to non-personal data. These include access to car maintenance data, truck parking data and energy network interoperability data.

We note also that, apart from market failures, data sharing obligations may **trigger equity concerns**. The GDPR contains restrictions on data processing to protect the personal data of the data subject, based on the objective of stimulating the free movement of personal data as well as the status of data protection as a fundamental right. A similar reasoning can be applied to non-personal data where the protection of commercial confidentiality is assumed to enhance private and social welfare. There may be cases, however, where private and social welfare diverge. A reduction in private rights to object to data re-use can increase social welfare, though not necessarily a strictly Pareto-welfare increase.

References

Acemoglu, D., A Makhdoumi, A Malekian and A Ozdaglar (2019) Too much data: prices and inefficiencies in data markets, NBER Working Paper 26296.

Acquisti A., C Taylor and L. Wagman (2017) Economics of privacy, Journal of Economic Literature 54(2), 442-492.

Agarwal A., J. Gans and A. Goldfarb (2018) Prediction machines, Harvard Business Review Press.

Atik C. and B. Martens (2020) Agricultural data markets and competition in agricultural services, a farmer perspective, forthcoming.

Baron J, J. Contreras, M. Husovec and P. Larouche (2019) Governance of standard development organisations, the case of intellectual property rights, European Commission Joint Research Centre, Science for Policy report.

Bergemann D. and A. Bonatti (2018) Markets for information: An introduction, CEPR Discussion Paper DP 13148.

Bergemann D., A. Bonatti, and T. Gan (2020) The Economics of Social Data Cowles Foundation Discussion Paper 2203.

Bertacchini, E, De Mot J and B Depoorter (2009) Never two without three: commons, anticommons and semicommons, Review of law and economics 5(1), 163-176.

Bourreau, M. and A. de Streel (2019) Digital conglomerates and EU competition policy, Working Paper.

Bresnahan T.F. and M. Trajtenberg (1995) 'General purpose technologies: Engines of growth?', Journal of Econometrics 65(1), 83-108.

Brynjolfsson, E, A Collis, E Diewert, F Eggers and K Fox (2019) Accounting for the Value of New and Free Goods in the Digital Economy, NBER Working Paper 5695.

Buchanan J. and Y. Yoon (2000) Symmetric tragedies: commons and anti-commons, Journal of Law and Economics 34(1), 1-13.

Caillaud B. and B Jullien (2003) Chicken & egg: Competition among Intermediation Service Providers, RAND Journal of Economics 34(2), 309-328.

Carballa-Smichowski Bruno (2018) Determinants of coopetition through data sharing in MaaS. Management & Data Science 2(3). https://doi.org/10.36863/mds.a.4160

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

Colangelo, G. and Maggiolino, M. (2018) Big data as misleading facilities, European Competition Journal, 249-281.

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

Ctrl-Shift (2018) Data Mobility: The personal data portability growth opportunity for the UK economy, Report for the UK Department for Digital, Culture, Media & Sport.

De Cornière A. and G. Taylor (2020) Data and Competition: A General Framework with Applications to Mergers, Market Structure, and Privacy Policy, CEPR Discussion Paper 14 446.

Deloitte et al. (2018) Emerging issues of data ownership, interoperability, (re-)usability and access to data and liability, Study for the European Commission.

Dengler, Sebastian and Prufer, Jens (2018) Consumers' Privacy Choices in the Era of Big Data. TILEC Discussion Paper 2018-014.

De los Santos, B. and S. Koulayev (2017) Optimizing Click-Through in Online Rankings with Endogenous Search Refinement, Marketing Science 36(4), 471-643.

De Streel A. and T. Tombal (2020) The Fifty Shades of Data Sharing and the Law, available at SSRN.

Dosis and Sand-Zantman (2019) The ownership of data, Mimeo.

Drexl J. (2017) Designing Competitive Markets for Industrial Data - Between Propertisation and Access, JIPITEC, 257-292.

Drexl J. (2018) Data access and control in the era of connected devices, Report for BEUC.

Duch-Brown N., F. Mueller-Langer and B. Martens (2016) The economics of ownership, access and trade in digital data. JRC Digital Economy Working Paper 2017-1.
Everis (2018) Data sharing between companies in Europe, Study for the European Commission.
European Commission (2017a) Communication Building a European Data Economy, COM(2017) 9.
European Commission (2017b) Staff Working Document Better Regulation Guidelines, SWD(2017) 350.
European Commission (2018a) Communication Towards a common European data space, COM(2018) 232.
European Commission (2018b) Commission Staff Working Document, Guidance on sharing private sector data in the European data economy, SWD(2018) 125.
European Commission (2020) Communication A European Strategy for Data, COM(2020) 66.
EU Support Centre for Data Sharing (2020), Analytical report on EU law applicable, to sharing of non-personal data.
Feasey R. and A. de Streel (2020) Data Sharing for Digital Market Contestability: Towards a Governance Framework, CERRE Report.
Fingelton and ODI (2019), Open Banking: preparing for lift off.
Franck J. U. and M. Peitz (2019) Market Definition and Market Power in the Platform Economy, CERRE, Report.
Furman J., D. Coyle, A. Fletcher, D. McAuley and P. Marsden (2019) Unlocking digital competition, Report of the digital competition expert panel.
Gal M.S. and D.L. Rubinfeld (2019) Data Standardization, NYU Law Review 94(4), 737-770.
Graef I. (2016) EU Competition Law, Data Protection and Online Platforms: Data as Essential Facility, Kluwer Law International.
Graef I., Husovec, M. and N. Purtova (2019) Data Portability and Data Control: Lessons for an Emerging Concept in EU Law, German Law Journal, 359-1398.
Graef I., Tombal, T. and A. de Streel (2019) Limits and Enablers of Data Sharing. An Analytical Framework for EU Competition, Data Protection and Consumer Law, TILEC Discussion Paper 2019-024.
Graef I., Husovec, M. and J. van den Boom (2020) Spill-overs in data governance: uncovering the uneasy relationship between the GDPR's right to data portability and EU sector-specific data access regimes, Journal of European Consumer and Market Law 9(1), 3-16.
Iansati M. and K. Lakhani (2020) Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World, Harvard Business Press.
IDC and Lisbon Council (2019) European Data Monitor, Study for the European Commission, available at: http://datalandscape.eu/
Jones C. and C. Tonetti (2019) Non-rivalry and the economics of data, NBER Working Paper 26 260.
Kerber W. (2016) Governance of Data: Exclusive Property vs. Access, IIC 47, 759-762.
Krämer J., D. Schnurr D. and S. Broughton-Micova (2020) The role of data for digital market contestability, CERRE Report.
Krämer J., P. Senellart and A. de Streel (2020) Making data portability more effective for the digital economy, CERRE Report.
Lerner A. (2014) The Role of 'Big Data' in Online Platform Competition, available at SSRN.
Lundqvist, B (2018), Competition and Data Pools, Journal of European Consumer and Market Law, 146-154.
Lynskey O. (2017) Aligning Data Protection Rights with Competition Law Remedies? The GDPR Right to Data Portability, European Law Review 42(6), 793-814.
Madison, M, K Strandburg and B Frischmann (2016) Knowledge commons, Univ of Pittsburg Legal studies research paper series 2016-28.
Martens B. and F. Müller-Langer (2020) Access to digital car data and competition in aftersales maintenance markets, Journal of Competition Law and Economics 16(1), 116-141.

Mayer-Schönberger V. and T. Ramge (2018) Reinventing Capitalism in the Age of Big Data, John Murray Press.
McNamee R. (2019) Zucked. Waking up to the Facebook catastrophe, Penguin Press.
Milgrom, D. North and B. Weingast (1990) The role of institutions in the revival of trade, Economics and politics 2(1), 1-23.
Montes R., W. Sand-Zantman and T. Valletti (2018) The value of personal information in online markets with endogenous privacy, Management Science 65(3), 1342-1362.
Motta M. (2004) Competition Policy: Theory and Practice, Cambridge University Press.
OECD (2013), Exploring the economics of personal data: a survey of methodologies for measuring monetary value, DSTI/ICCP/REG(2011)2.
OECD (2015) Data-Driven Innovation: Big Data for Growth and Well-Being, OECD Publishing.
OECD (2016) Maximising the economic and social value of data: Understanding the Benefits and Challenges of Enhanced Data Access, OECD Directorate for Science, Technology and Innovation, paper DSTI/CDEP (2016) 4.
OECD (2019) Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies, OECD Publishing.
Panzar J. and R. D. Willig (1981) Economies of Scope, American Economic Review 71(2), 268-272.
Palfrey J. and U. Grasser (2012) Interop: The Promise and Perils of Highly Interconnected Systems, Basic Books.
Parker G. and M. Van Alstyne (2017) Innovation, Openness, and Platform Control, Management Science, 64(7), 1-18.
Parker G. and M. Van Alstyne (2005) Two-Sided Network Effects: A Theory of Information Product Design, Management Science 51(10), 1494-1504.
Powles J. and H. Hodson (2017) Google DeepMind and healthcare in an age of algorithms, Health and Technology 7, 351–367.
Prüfer J. and C. Schottmüller (2017) Competing with Big Data, TILEC Discussion Paper 2017-006.
Prüfer J. (2020) Competition Policy and Data Sharing on Data-driven Markets: Steps Towards Legal Implementation, available at http://library.fes.de/pdf-files/fes/15999.pdf.
Richter H. and P. Slowinski (2019) The Data Sharing Economy: On the Emergence of New Intermediaries, IIC 50(4), 4-29.
Rochet J.C. and J Tirole (2003) Platform Competition in Two-sided Markets, Journal of the European Economic Association, 1(4), 990-1029.
Rochet J.C. and J Tirole (2006) Two-sided markets: a Progress Report, Rand Journal of Economics, 37(3), 645-667.
Rosen S. (1983) Specialization and human capital, Journal of Labor Economics 1(1), 43-49.
Schultze N., F. Parisi and B. Depoorter (2002), Fragmentation in property: Towards a general model, Journal of Institutional and Theoretical Economics 158, 594–613.
Schweitzer H., Haucap J., Kerber W. and Welker R. (2018), Modernising the law on abuse of market power, Report for the German Federal Ministry for Economic Affairs and Energy.
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.
Smith, H. (2005) Governing the tele-semi-commons, Yale Journal of Regulation 22, 289-314.
Teece D. (1980) Economies of scope and the scope of the enterprise, Journal of economic behaviour and organisation, 223-247.
Teece D. (1982) Towards an economic theory of the multi-product firm, Journal of economic behaviour and organisation, 39-63.
Tombal T. (2020) Economic dependence and data access, IIC 51(1), 70-98.
Soto I. and M. Gomez-Barbéro (2019) The contribution of precision agriculture technologies to farm productivity and the mitigation of greenhouse gas emissions in the EU, Technical Report, Joint Research Centre of the European Commission.

UK CMA (2020) Online platforms and digital advertising: Market study Final Report.
VVA (2017), Data in platform-to-business relations, Study for the European Commission.
Weibe, A. (2017) Protection of industrial data - a new property right for the digital economy?, Journal of Intellectual Property Law & Practice 12(1), 62-71.
Williamson O. (1985) The economic institutions of capitalism, Simon and Schuster.
Zech, H. (2016) "Data as tradeable commodity", in De Franceschi, A. (ed.) European Contract Law and the Digital Single Market, Intersentia, 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.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from EU Bookshop at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

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.

EU Science Hub

ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub