Personal Data Portability in the Platform Economy: Economic Implications and Policy Recommendations[1]

Forthcoming at *Journal of Competition Law & Economics*

Jan Krämer[*]

University of Passau, Chair of Internet and Telecommunications Business,
Dr.-Hans-Kapfinger-Str. 12, 94032 Passau, Germany

and

Centre on Regulation in Europe (CERRE)
Av. Louise, 475 (box 10), 1050 Brussels, Belgium

Abstract

Article 20 of the General Data Protection Regulation (GDPR) gave consumers in the European Union the right to port their personal data between digital service providers. We critically assess the economic implications of this new right in the light of the extant economic literature and with a focus on competition and innovation in the digital platform economy. In particular, we conclude that observed user behaviour data should clearly fall under the scope of data portability and that, above and beyond the regulations set out under GDPR, a right to port personal data continuously and in real-time would be necessary to truly empower consumers in the context of the digital platform economy. We also discuss the economics of Personal Information Management Systems (PIMS), which many policymakers see as an essential tool for consumers in an economy where data portability becomes more widespread. However, we are sceptical that PIMS will be self-sustainable and instead advocate to facilitate the development of open-source projects, which have made little progress so far due to a lack of interfaces (which would come about with a right to continuous data portability) and due to a lack of common standards.

Keywords: Data Portability, Personal Information Management Systems, GDPR, Data Economy, Internet regulation

JEL codes: K2, L51, L96

[1] The author is grateful to (in alphabetical order) Malte Beyer-Katzenberger, Marc Bourreau, Richard Feasey, Claire-Marie Healy, Bertin Martens, Pierre Senellart, Alexandre de Streel, Thomas Tombal, as well as two anonymous reviewers for very helpful comments and suggestions. The paper is largely based on Sections 4, 5 & 6 in Krämer, Senellart and de Streel (2020). Making Data Portability More Effective for the Digital Economy. CERRE Policy Report. Available at: https://www.cerre.eu/publications/report-making-data-portability-more-effective-digital-economy

[*] Tel: +49 851 509 2580, fax: +49 851 509 2582, e-mail: jan.kraemer@uni-passau.de. The author is a full professor at the University of Passau and an Academic Co-Director of the Centre on Regulation in Europe.

1 Introduction and background

After a long political tug-of-war the new European General Data Protection Regulation (GDPR) came into force on 25 May 2018 (European Commission, 2016b). It replaces the previously applicable European Data Protection Directive (95/46) from 1995, which originated from a time when personal data did not have the economic and societal importance that it has today, where massive amounts of personal data are collected, analysed, and monetised every day, particularly in the digital economy – making use of advanced (big) data analytics.

From a legal and economic point of view, there are important elements in the GDPR that strengthen the rights of data subjects. Of particular note here is the "right to data portability" under Article 20 GDPR, which is intended to enable users to exercise more control over their personal data. It is also supposed to enable users to counteract lock-in effects in digital services and facilitate switching to an alternative content or service provider.

However, the GDPR was built on the premises of fundamental rights but not on the premises of economic regulation or competition law, although the right to data portability may have competition-enhancing effects. As we will discuss in this article, to date very little research exists on the actual economic impacts that come about with the right to personal data portability, and indeed, the extant economic literature shows that complex economic trade-offs can arise from data portability.

Furthermore, we critically assess whether additional legal requirements and clarifications are needed to make the portability right more effective in the specific context of the digital platform economy. For example, the right to data portability under Article 20 GDPR does not encompass a continuous porting of data (facilitated, e.g., by Application Programming Interfaces or APIs), and it is also currently not entirely clear to what extent users can port their observed data (i.e., data implicitly given, e.g., through clicks and location). In this context, we argue that continuous, real-time access to users' volunteered (i.e., data explicitly given) *and* observed data may be crucial to stimulate competition and innovation in the digital economy. However, this is an issue that is controversially debated. Some note that with frictionless data portability, it is to be feared that the quality of content and service offers (e.g., financed by advertising) will decline because, with data portability, more companies will have access to the same personal data, which may well intensify the competitive situation on the data and advertising market. In other words, third parties could act as free-riders on the data market, which could ultimately also harm the customer. In reverse, others note that the

free flow of personal data (with appropriate safeguards to mitigate privacy and security risks, and with the users' consent) would stimulate innovation activities because data would be freed from its silos, and a much larger set of minds could get access to it. More generally, the OECD (2019) noted that data access and sharing is estimated to generate social and economic benefits worth between 0.1% and 1.5% GDP in the case of public-sector data, and between 1% and 2.5% of GDP (in a few studies up to 4% of GDP) when also including private-sector data.

We also comment on the economics and the economic viability of Personal Information Management Systems (PIMS), who are believed to play a central role in a digital economy in which data portability becomes more widespread. Without the help of PIMS, consumer may struggle to deal with the complexities and processes that come about with data portability, and which require time and expertise. As such, data portability is only ever going to work effectively if the process is both transparent and easy from the consumer's perspective. Among other things, PIMS are supposed to offer users a centralised dashboard to monitor and control the flow of their personal data, and may even empower consumers to participate from the economic value of their personal data. We offer a critical assessment in this regard and highlight that, despite of their importance in the context of data portability, it is not likely that PIMS will find a sustainable business model, especially if this is based on selling personal data.

Finally, in view of the digital platform economy, we make a number of policy recommendations on *how personal data portability can be made more effective* in order to better empower consumers with respect to the use of their data and the switching and multi-homing of digital services.

2 Preliminaries: The economics of data and some terminology

Understanding the economics of data portability requires foremost to acknowledge the economics of data. Here we focus at first on the differentiation between data, information and knowledge, only from the latter of which ultimately value can be derived. Moreover, we take a closer look at the economic boundary conditions of the non-rivalry of data.

2.1 The value of data, information and knowledge

Data per se does not have any economic value as it is merely the (digital) representation of signals that have been received or perceived using some syntax. For example, the receipt of a light signal can be transformed into data by recording the time (e.g., using the syntax HH:MM:SS) and recording the "on" and "off" states (e.g., using the syntax "1" for "on" and "0" for "off"). Such data is transformed into information only if it is combined with semantics. For example, a corresponding semantic would be that the data is on received light as part of a communication effort using Morse code. This gives the data a meaning (here: a message that is communicated), hence transforming it into information. Such information can then be transformed further into actionable knowledge with the additional input of and in combination with other pieces of information. For example, the received message may have been "HELP" and combined with the information that a friend is hiking all by himself in the mountains, in about the location from where the light beams have been received, leads to the actionable knowledge that he is in danger and that a rescue operation should be started. The same holds true, for example, for clicks (data) on an e-commerce site, which represent which products a shopper considered for purchasing (information) and which can then be used to infer which products the shopper might be interested in (knowledge). Ultimately only such actionable knowledge that is generated from data potentially has economic value and can increase welfare.

Nevertheless, it is customary to refer to all three concepts – data, information, and knowledge – only as "data" in policy circles. Instead, 'raw data' is often differentiated from 'derived' and 'inferred data' (see Section 2.2). We will, therefore, follow this practice as well in the following. To do so, we need to introduce a related but distinct terminology in the next subsection.

2.2 Volunteered, observed, and inferred data, and their relation to data portability

It is also important to note a differentiation in how data was acquired about an individual consumer. As is customary, we distinguish between volunteered, observed and inferred data.

Volunteered data is explicitly and intentionally revealed by a user, such as name and date of birth entered into a registration form, a post, tweet or rating submitted, or an image or video uploaded. Consumers are usually aware of the volunteered data that they revealed and often this is the only type of data that consumers think they have revealed when using an online service. In practice, it is often also the only data that is fully made available by data controllers in response to a data portability request according to Article 20 GDPR.

Observed data is obtained from the usage of a device, website or service and the user may or may not be aware that such data is collected. This ranges from clicks on products and purchase histories over geo-locations gathered by GPS sensors in smart phones to recording every single interaction of the consumer with the service–potentially even when the consumer does not even know that she is currently interacting, such as in the context of voice assistants that are constantly recording. As we will discuss below, there is some uncertainty regarding the degree and scope to which observed data should be fully made available according to Article 20 GDPR.

Inferred data is derived through refinement and recombination from volunteered and observed data, e.g., by use of data analytics such as clustering, filtering or prediction. The result can be a complex preference profile of a consumer or a recommendation. Inferred data can actually already be – using the definition introduced in the previous subsection – knowledge that can provide actionable insights. Thus, inferred data is ultimately the basis for competition between data-intensive firms, whereas volunteered data and observed data are the 'raw data' inputs.

The distinction between volunteered, observed and inferred data, albeit not being a legal definition, is also important in the context of the scope of data portability. Article 20 GDPR notes that only data that is "provided" by the data subject is covered by the data portability right. Moreover, only personal data whose processing is based on consent or a contract is subject to Article 20 GDPR. In its interpretative Guidelines, the European Data Protection Board (EDPB) clearly includes volunteered data in the scope of Article 20 GDPR, while inferred data should not be included (European Commission 2017a, p.10). However, some legal uncertainty arises with respect to observed data. The EDPB notes that the portability right should be interpreted broadly and should cover observed data as well. If this interpretation is followed, for example web tracking data and clickstream should be covered by the portability right as this is observed data. However, it remains to be seen whether future case law will follow such a broad interpretation.

2.3 Non-rivalry of data, and its limits

Data is *non-rival*, which means that the same data can, in principle, be used by different entities at the same time. The same data could also be shared and collected by different entities without depleting the source of data for others. For example, many observers could have collected the data on the light signals sent at the same time without interfering with the ability of others to do the same. Likewise, the data on the light signals could have been shared without having to give it up.

But data is also *excludable*, which means that the data controller can impose technical or legal constraints to prevent sharing of data. Non-rivalry and excludability are distinct concepts and should not be seen as the two sides of the same medal. Although the consumption of data is non-rival, the collection of data as well as the derivation of value from data are subject to intense competition, i.e., the source of "rivalry" between firms. In the following, we therefore put non-rivalry of data (as opposed to the excludability of data) in perspective of such rivalry for data between firms, in order to emphasise benefits and risks of data sharing and data portability.

2.3.1 Competition in the collection of data

Specific data (e.g., on products liked on a particular e-commerce site, or links clicked on a particular search engine*) cannot just be collected by anyone interested*. Like a focused ray of light, e.g. emitted by a laser, cannot by received by a random observer but just in a particular location. In this sense, although data consumption of data is non-rival, *there can be intense competition in the collection of data*, and we discuss this in more detail below.

First, we note that there is a lively debate with respect to the degree to which the collection of data is contested. On the hand, some scholars claim that data is ubiquitous, as consumers are willing to share their data over and over again with different services, frequently multi-home similar services, and that specialised data brokers make data available to everyone who wants to buy it (see, for example, Lambrecht and Tucker 2015, and Tucker 2019). On the other hand, this is contrasted by the empirical findings that – despite the multitude and variety of websites and online services available – consumers' attention is highly concentrated on a few sites and even fewer firms. In other words, only those firms are in the right 'location' to actually collect consumer data at a large scale. For example, the European Commission found in the context of the Google AdSense case that Google had a market share of generally over 90% in 2016 the market for general search in all Member States (European Commission, 2019b). Similarly, in its investigation of Facebook, the German Federal Cartel Office found that Facebook had market share in the market for social networks of over 95% (with respect to daily active users) in Germany in December 2018 (Bundeskartellamt, 2019). In similar vein, even fewer firms are currently able to collect tracking data across multiple sites. For example, Englehardt and Narayanan (2016) measured which third-party web trackers were deployed at the top 1 million websites. They find that Alphabet/Google (with trackers deployed at about 70% of all sites), followed by Facebook (trackers deployed at about 30% of all sites), are also in a unique position to track users' activity across various (third-party) websites. Very similar results are

obtained by Ghostery (Macbeth, 2017)[2], a browser extension that blocks third party trackers. The situation is likely to become even more pronounced as Google has recently announced to disallow third-party cookies in Google's Chrome browser, which many view as a step that bolsters Google's and Facebook's dominance in web tracking (Barker, 2020), because these companies have alternative means to track users across the web, e.g., through services such as 'Google Analytics' or 'Login with Facebook'.

Taken together, this already points to the conclusion that the collection of observed user data (as opposed to volunteered user data) is indeed often highly concentrated, because for key services (such as search, or social networking) the market is highly concentrated and only a few firms are able to track user activity across the web. This conclusion is likely to be strengthened as Internet-enabled smart devices (such as smartphones, smart speakers, TVs, cameras and other Internet-of-things devices in the households) become more and more prevalent. These devices have the potential to track consumers' behaviour and daily activities also when they are not surfing the web and are supposingly 'offline'.

Thus, observed consumer data is not ubiquitously available, and it is also usually not feasible nor socially desirable to duplicate the collection of the same observed data. This would mean that users would have to conduct the same search, the same post or the same purchase on several platforms; it would mean that even more web trackers are being built into the websites that we visit; and that we would need to invite smart devices from even more firms in our homes. Thus, concentration in the collection of data is not necessarily a problem, but it does provide a strong rationale for sharing data.

2.3.2 Competition in deriving value from data

The economic value of data likely depends on how many others have access to the same data, or put more precisely, can derive the same insights from data. For example, both Ishihashi (2019) and Gu et al. (2018) highlight by means of a game-theoretic model that the value of data collected from consumers may drop significantly (in their theoretical models to zero) if more than one firm possesses it, i.e., if data is non-exclusive. Once the data is created (e.g. generated by using the service of a firm, which has 'paid' for the data by offering a free service), consumers will give it up

[2] See also https://www.ghostery.com/study/

to a second firm, even at a 'price' close to zero, because each additional sharing of data does not bear opportunity costs. This is a direct consequence of the non-rivalry of data. This means that if data sharing is frictionless and bears zero transaction costs for consumers, firms eventually possess identical sets of data. A potential buyer of this data is only interested in acquiring such data once because each data set is a perfect substitute for the other. This means that firms engage in fierce price competition selling the data – known as Bertrand competition. Eventually they compete each other down to marginal costs, which means that they sell the data for a price close to zero. If, however, one firm would have possessed the data exclusively, it could have demanded a non-zero price. In this sense, although the consumption of data is non-rival, the *economic value that can be derived from data is contested*.

If taken literally, this provides a strong rational for *not sharing* data, as this would destroy any incentive to collect data in the first place. However, three important clarifications are in order. First, and foremost, the above argument does not differentiate between 'data' and 'knowledge', because it essentially only considers data intermediaries, which collect and sell raw data. Even though two firms may have access to the same raw data set (in terms of volunteered and observed data), they may derive different insights from it ('inferred data' or 'knowledge' in our terminology), which is ultimately the basis for competition.

Second, and relatedly, the above discussion has abstracted from cases where the data is not sold to third parties on some data market, but rather used internally (e.g. for marketing purposes or for improving the service quality)—or where data is combined with other data available to the firm and the enriched data set can be sold as a unique data set, overcoming the competition in the data market.

Third, the above argumentation has abstracted from transaction costs, such as additional privacy concerns of sharing the data set with another firm, or the effort in selling additional data in return for only a low additional benefit. If these transaction costs are non-negligible, this will reduce the non-rival nature of data sharing, which leads to less data sharing and eventually decreases the competition in the data market.

Taken together, this means that more prevalent sharing of 'raw' user data will likely render the market for data intermediaries (which simply acquire and sell raw data, but do not offer advanced analytics on such 'raw' data) more competitive and possibly unlucrative. However, this does not

destroy the incentives to compete on the basis of insights derived from data. Rather, as raw data becomes more prevalent, the focus of competition is likely to move more from collection to analytics, which likely stimulates innovation rather than stifling it (see Section 4.2). Indeed, as data collection is inherently concentrated (see Section 2.3.1) and the services through which (observed) data is collected usually exhibit strong network effects (see Section 3.2), a stronger competition at the data analytics level (i.e., based on knowledge) seems much more feasible and desirable than competition at the data collection level.

2.4 The quality of data, and its relationship to volunteered and observed data

Volunteered, observed and inferred data are also useful concepts for discussing different qualities of data. Generally, the *quality of data* can be measured along the dimensions of

- fitness for use (is the data suited to derive the desired insights?)
- accuracy (does the data represent the facts?)
- completeness (how many data points are missing?)
- timeliness (how fast can data be collected and how quickly is it outdated?).

Volunteered data is derived from direct human input. That is, this data may be inaccurate, e.g., because wrong information (e.g. a wrong email address, fake name or fake review) have been submitted intentionally or unintentionally. But often the accuracy of the volunteered data is also essential for the quality of the service, which provides consumers with an incentive to provide accurate data (e.g., a correct liking of songs in a music streaming service will trigger a better recommendation for new songs). Moreover, volunteered data is prone to being incomplete, and it may outdate relatively fast, because it is not automatically updated after it has been provided. However, volunteered data is usually structured, because it has been collected in a structured way, such as through forms, 'like' buttons, or on a rating scale. Thus, it can immediately be used as input to generate inferred data.

Observed data is less prone to deliberate manipulation, because it is derived from actual behaviour and sensors. Moreover, observed data tends to be more complete and timelier, because it is recorded automatically. The accuracy and fitness for use is often very context dependent. For example, click data from an e-commerce session can be very noisy and sparse, because the user might just be browsing through random products and in each product category only very few products are explored. In another session, the similar click data can be very accurate and dense, as a consumer

explores several similar products and puts some of them in the shopping basket, but finally only buys one. Similarly, data from sensors (e.g., GPS sensors) can be highly accurate at times and inaccurate at other times, depending, for example, on geography and environmental conditions. Quality of data with respect to fitness for use also depends highly on the context. Highly accurate GPS data, for example, may be necessary to identify which products a consumer was interested in when visiting a department store, whereas coarser data may still be acceptable to identify which stores a consumer has visited in a mall. In any case, observed data is often less structured and must be cleaned and structured in a way that allows to derive actionable knowledge.

Finally, the quality of inferred data depends not only on the quality of the observed and volunteered data, but also on the amount of observed and volunteered data. With respect to the analysis of data, empirical studies suggest that in many (big) data analytics applications,

- there is a minimum required scale,
- there are benefits from larger data sets, and
- these benefits are marginally decreasing as data sets become very large.

More precisely, Junqué de Fortuny et al. (2013) and Martens et al. (2016) demonstrate that prediction accuracy increases for larger data sets of fine-grained user behaviour data (observed data). Whereas benefits decrease marginally as prediction accuracy approaches the theoretical benchmark (cf. Li, Ling, Wang, 2016), the studies show this convergence is not yet reached in many popular application settings. Furthermore, for the online advertising industry, Lewis and Rao (2015) find that only very large amounts of data allow firms to measure whether advertising campaigns are indeed successful. Thus, empirical studies and general indications point to the presence of scale economies from data collection and data analysis.

Consequently, having access to more data (e.g., not only volunteered but also observed data) will, in many applications, yield a better quality of the inferred data (i.e., the actionable knowledge) and thus offer higher profit opportunities for firms. This highlights again that the scope of data portability, i.e., whether it is restricted to volunteered data or also encompasses observed data, is crucial from an economic perspective.

3 Data portability and competition

In the previous section we derived that (i) particularly observed data is a valuable raw input for data-intensive business models in the digital economy, and (ii) the collection of observed data in the digital economy is inherently concentrated and only a few digital firms are in a unique position to collect it. In this context, the question arises how newcomers and start-ups may get access to the required observed data, in order to be able to compete on the basis of inferred data, i.e., knowledge and insights generated from these raw inputs. More generally, this raises the question how and if data portability indeed increases the competitiveness of digital markets.

In doing so, we take the perspective of a consumer, and highlight that switching to a new service may impose two types of 'data costs' that can result in consumer lock-in. The first type of cost is a transaction cost from switching. The second type of cost is related to network effects. We describe both in turn.

3.1 Data portability and data-induced switching costs

It is often argued that consumers do not switch from one digital service to another because they shy away from the transaction costs to give away their (volunteered) data again at the new service. This seems especially problematic in cases where large amounts of data have been volunteered over a long time in which the current service was used. For example, in the case where thousands of songs have been liked while using an online streaming service, liking the same songs again at a new service seems an unreasonable burden. This transaction cost is a classic switching cost, i.e., a fixed cost for setting up a service that occurs only once. When a consumer evaluates two services—the one that she is currently using, and the new one—the difference in expected utility must at least exceed the switching cost, in order for the consumers to switch.

The classic literature on switching costs (see, e.g., Klemperer 1987a) finds that switching costs can constitute a significant barrier to entry, shielding incumbents from competition. In digital markets switching cost may vary substantially depending on the context. However, the classic literature also finds that when established services compete for customers in the presence of switching costs, then competition is fierce for 'new' customers, whereas 'old' customers tend to be exploited (see, e.g., Klemperer 1987b; Farrell and Shapiro 1988). However, in the long run, markets tend to be less competitive in the presence of switching costs (see, e.g., Beggs and Klemperer, 1992).

Generally, services whose quality depend to a high degree on customisation and personalisation (e.g., services in which recommendations play a significant role) are more prone to be subject to switching costs. However, often it may not just be volunteered data that constitutes a switching cost, but also the observed data. For example, the current music streaming service may also have recorded which songs were actually listened to, how often each song was played, for how long, and at what time of the day. Like the volunteered data, this observed data can be a very useful input for the next music streaming service.

The right to data portability can lower these switching costs by making the volunteered data and observed data readily available in a "structured, commonly used and machine-readable format" (Article 20, GDPR) to the consumer, who can then pass it on to the new provider. Thus, in light of the classic switching cost literature, the right to data portability can make digital markets more competitive in the long run and lower entry barriers for new service providers. This is commonly viewed as beneficial to consumer welfare and one of the strongest economic arguments for the right to data portability.

However, more recently a new strand of economic literature has re-investigated the classic results and specifically considered the welfare implications of the right to data portability. In a game-theoretic model, Wohlfarth (2019) shows that the right to data portability can have an effect on the amount of data this is collected by data-intensive firms. Without the right to data portability, market entrants are forced to design services that economise on the use of data in order to be able to attract consumers. However, as data can be easily ported to the entrant, the new provider has less incentives to economise on data use and increases the amount of data collected. In this sense, the GDPR's right to data portability (Article 20) runs contrary to the GDPR's principle of data minimisation (Article 5.1c); not only from a legal point of view, but also with respect to the economic incentives of data collection. Wohlfarth shows that this economic trade-off can eventually lead to a reduction in consumer surplus.

In a similar vein, Krämer and Stüdlein (2019) also analyse the economic effects of data portability on market entry in a game-theoretic model. They focus on the firms' incentives to disclose user data, e.g., in the context of targeted advertising, with and without the right to data portability. They show that the right to data portability is likely to benefit the 'old' customers of the incumbent, especially those that do switch to the new provider, as switching costs are reduced and competition is increases. However, the 'new' customers of the entrant, i.e., those that were not previously customers of the

incumbent, are likely to be worse off, because the entrant's competitive position is strengthened under the new right to data portability. Without data portability, the entrant would have competed more fiercely for these new customers. In reverse, this means that its customers are worse off than without data portability. Again, this highlights that not all consumers need to benefit from a right to data portability – although this right unambiguously lowers switching costs.

Despite these nuances, if data portability indeed lowers switching costs, this is likely to increase the competitiveness of markets. However, not the least, this will also depend on whether consumers actually make use of data portability, and whether it is possible to imported this data at other services.

3.2 Data portability and network effects

Network effects arise whenever a consumer's value of a good or service depends on how many other consumers are using the same good or service. Network effects are ubiquitous in digital markets, and often services are explicitly designed to incorporate network effects. For example, in social networks, network effects arise, because participation in the network is more valuable the more other people are also using the same social network. This is a direct network effect. But more than often indirect network effects are also present. In this case the value of the service increases because of the presence of more complementors to the service. For example, an operating system is valuable mostly due to the availability of software complements that run on this operating system. Likewise, an e-commerce website may be valuable to a consumer due to the number of product reviews on that website, which depend only indirectly on the number of users. Indirect network effects are also at the core of platform markets (multi-sided markets), which bring together at least two distinct user groups (such as buyers and sellers). At least one of the groups values the presence of another group on the platform, thereby creating an indirect network effect. A prototypical example is an app store, where consumers value the presence of many app developers, and, in reverse, app developers value the presence of many consumers. Network effects are important in the context of data portability and the competitiveness of markets for two main reasons, which we discuss in the next two subsections.

3.2.1 Data portability and user-side network effects

Network effects create a coordination problem. Because the value of the service depends directly or indirectly on how many others are using it, consumers want to be where everybody else is. This also

creates a lock-in situation, distinct from that of simple switching costs, because switching a provider seems only reasonable if everyone switches at the same time. It is important to note that, contrary to the case of data-induced switching costs, data portability does not alleviate this type of lock-in. This would require some (protocol) interoperability (see Crémer et al. 2019) of the services, whereby services interoperate to a degree where ultimately users can interact seamlessly albeit being on different networks – like users of different telecom networks can communicate with each other. Then users can switch to a new provider without losing access to the network effect exerted by users who remain with the old provider. Consider a social network for example. Even if a user would be able to take its data to a new social network, it would still not be able to interact with the users that remained on the old network. Indeed, in this context, it has been argued that "identity portability" (Gans, 2018) or "social graph portability" (Zingales and Rolnik, 2017)—both a form of protocol *inter*operability—would be desirable to overcome user-side network effects. Identity portability means that a person can switch to a new network and take her identity with her, so that all messages related to that person are forwarded to the new network, and vice versa. The idea of identity portability is thus comparable to interconnection in conjunction with number portability on telecom networks.

However, demanding (protocol) interoperability over and beyond (data) portability also has some caveats, especially with respect to the need for regulatory oversight (like in telecoms networks), and the ensuing risk of barriers to innovation due to the necessity to remain within the standard for interoperability. As others have noted (see, e.g., Crémer et al. 2019), this seems warranted only in specific applications such as text messaging services and social networks, where the benefits of interoperability (through increasing the network effect and competition *in* the market) are likely to outweigh the risk (of reduced innovation and competition *for* the market).

There is also a noteworthy interaction between network effects and switching costs, laid out in Suleymanova & Wey (2011). Markets with strong network effects tend to monopolise, because consumers tend to gravitate to the service or platform that already exhibits the largest network effects. In other words, once a critical mass of users has been reached, markets tip towards the largest player. Switching costs can dampen this process, because they create an economic friction (transaction cost) that prevents customers from switching to the service with higher network effects as easily. In this vein, switching costs may allow two networks to co-exist at the same time. However, this is usually not an efficient situation in the presence of network effects. Moreover, the argument

rests on the assumption that there are two services, albeit with different market shares, which both have a viable and stable user base. In practice, many digital markets with strong network effects have already tipped and new entrants do not have a viable and stable user base so that switching costs (or non-portability) would protect them from churn. Thus, we argue that in many relevant scenarios the interaction of data portability and network effects is not anti-competitive. But as laid out above, it is also not pro-competitive in the sense that data portability affects user-side network effects per se. Rather, data portability may impact analytics-based network effects, which may then have a pro-competitive effect. We describe this in the following.

3.2.2 Data portability and analytics-based network effects

Indirect network effects can also arise with respect to data analytics capabilities. Here network effects yield a positive feedback loop for algorithmic learning that can constitute an effective entry barrier (see Lerner 2014 for a thorough discussion): The more consumers are using a service, the more (volunteered and observed) data is created on which analytics can be performed and algorithms can be trained, which in turn results in an improvement of the service (e.g., better recommendations, better search results), which in turn leads to more consumers. For example, a dominant search engine is likely to provide better results simply because it records more search queries (volunteered data) and records more clicks on search results (observed data), which can then be used to derive better results lists for future searches.

This means, in this case barriers to entry are not created by switching costs in the narrow sense (indeed switching a search engine hardly entails any switching costs due to setting up the service), nor are they due to a lack of access to the network of users (on the same or another market side). Here it is rather the lack of access to the data that is created by fellow users – a type of indirect network effect – that creates a barrier to entry. This lack of data limits the ability of a new service provider to compete on the basis of algorithmic insights and data analytics, i.e., on the basis of inferred data or knowledge. This argument is explored more formally, for example, in Hagiu and Wright (2020), who show that this competitive advantage of the incumbent prevails under various assumptions about the shape of the learning curve from data. Moreover, Schaefer et al. (2018) provide empirical evidence that such network effects in algorithmic learning exist in the context of search engines.

Thus, if enough users would consent to a transfer of their raw data, and if it were possible to continuously transfer data through a standardised interface (API), then data portability could potentially promote entry and competition. It is important to highlight that the provision of data to competitors would be initiated by a specific consumer and, in each case, only entail the data of that consumer. This is very different to an access request entailing (anonymised) input data across a large number of users, initiated by another firm, e.g., by a competitor under the essential facilities doctrine. Although some commentators note that such access to input data may be a possibility to restore market contestability (e.g., by Argenton and Prüfer 2012, Krämer and Wohlfarth 2018 and Schweitzer et al. 2018), the focus of the present article is on user-initiated data portability. The advantage of data portability is that also personally identifiable data can be transferred, and thus there is no trade-off between competition and privacy regulation, which is inherent to access requests that are not user-initiated.[3] However, at the same time, it is unlikely that all users initiate a transfer of their data. Thus, the data set that is ported under data portability is likely to be more detailed on specific data subjects, but less representative for the user base as a whole. Whether or not such a data set is useful for a competing or complementing firm, is context specific and depends on the degree to which consumers make use of data portability, of course.

Finally, it is noteworthy to mention in this context that data portability may also be viewed with caution, because this can lead to situations in which ultimately consumers and competitors are worse off. In particular, Lam and Liu (2020) argue by means of a game-theoretic model that the right to data portability encourages consumers to reveal more data to the incumbent, because consumer are less concerned about data-induced switching costs that may arise later when considering to switch to a new market entrant (see Section 3.1). However, as consumers reveal more data, they also create a higher data analytics network effect at the incumbent, which indeed strengthens the competitive position of the incumbent vis-à-vis a new market entrant, and raises entry barriers. While data portability facilitates switching (which lowers entry barriers and raises consumers' surplus), this effect can be completely offset by the increase in the data analytics network effect (which raises entry barriers and may prevent efficient entry). In summary, the authors therefore

[3] However, even if data portability is in line with privacy regulation, data portability can still be the source for additional privacy risks. For example, as personal data is spread among more data controllers, there is a higher risk that it may be illegally accessed or exploited at one of them. We briefly return to this discussion in Section 6.3.2.

conclude that data portability can have an adverse effect on entry and long-run efficiency, although (or indeed because) data portability lowers switching costs. Note that this arguments rests strongly on the assumption that data portability leads to a different data revelation behaviour of consumers at the incumbent.

4 Data portability and innovation incentives

The previous section has focused on the impact of data portability on competition and contestability of markets, i.e., adopted a more static efficiency perspective. We now turn to a dynamic efficiency perspective and consider the impact of data portability, and more generally of data access on innovation incentives.

There has been a lively scholarly and policy debate about data access and innovation (see, e.g., Crémer et al 2019; Furman et al. 2018), which we do not intend to repeat here. However, we wish to highlight the main trade-offs involved in order to lay the groundwork for our policy recommendations in the context of data portability.

With regard to innovation, it is important to differentiate between the innovation incentives and capabilities of the firm that provides access to data and the firms that receive access to data. Moreover, it is important to differentiate whether such data is used to compete with the data provider or whether it is used for other purposes, such as offering complementary or completely new services. We consider these scenarios in turn.

4.1 Innovation by the incumbent: Conventional wisdom and kill zones

Although the consumption of data is non-rival (although there may be competition in the collection and monetisation of data, see Section 2.3), data is excludable, which – in an economic sense – means that a firm can exert exclusion rights on data assets. Without mandated access to data, data-intensive firms can utilise their economic control over data in order to make economic profits – be it by selling access to data or by using the data to improve their product or service in order to gain a competitive advantage. It is, by now, evident that data-rich firms can be highly profitable and this creates an economic incentive to invest in data collection and analysis. This spurs innovation, ranging from innovative services (that allow for a collection of data) to innovative data storage and data analytics. In this view, losing control of those data would lead to what economists call a "hold-up problem". That is, the lack of sufficient appropriability on data renders the economic benefits of data

uncertain and leads to a reduction in investment and innovation. This is, of course, conventional wisdom among economists, the very reason why intellectual property rights exist (i.e., a legal instrument for data *excludability*), and an argument that is not specific to data. In this sense, innovation incentives in the context of data are particularly strong when data can be used exclusively, and if in consequence a market can be monopolised.

In a similar vein, it is conventional wisdom in economics that there is a (non-linear) relationship between innovation incentives and competition, although there is continued research on the topic. Innovation is a means to provide a better service or product and to differentiate from competitors. This tends to increase profits and provides innovation incentives. In line with an Arrowian view, in a monopolistic environment, where high entry barriers already exist (be it by network effects or switching costs, or something else), innovation incentives tend to be low, because there is no competitive advantage to be gained from innovation. But in line with a Schumpeterian view, in markets with very high degrees of competition, innovation incentives also tend to be low as well, because innovation rents are quickly competed away and firms are often lacking sufficient scale for innovation activities.

Taking both arguments together, and in accordance with ample empirical evidence, innovation incentives tend to be the highest in oligopolies with only a few firms (see, e.g., Aghion et al. 2005). In this sense, if data portability indeed induces more competition in digital markets with high data-induced entry barriers, then this would likely increase incentives to innovate. In particular, in the context of digital markets, innovation incentives are particularly high if a market has not yet tipped and there is still competition *for* the market; or, possibly even more importantly, if digital markets were indeed contestable. This would mean that, despite a de-facto monopoly, entry barriers remain low and the incumbent needs to constantly defend its incumbency through innovation.

There is some doubt, however, as to whether the market inhabited by some big tech firms are indeed contestable and whether data portability would indeed lead to more competition in established markets. On the one hand, we have already detailed that data portability cannot overcome user-induced network effects per se (see Section 3.2), such that important barriers to entry remain, irrespective of the degree of data portability, if a new service were to compete head-to-head. On the other hand, there is growing empirical evidence that some firms may have established 'kill-zones' around their core business model (see, e.g., Kamepalli et al. 2020 and Scott Morton et al. 2019 for a thorough discussion, but also related news reports by The Economist 2018 and McLeod 2020). This

means that innovative start-ups, which may become competitors to a big tech firm's data-centric business model, may either be bought by the big tech firm, or the innovation is quickly integrated into the big tech firm's own service. In the latter case the incumbent has a comparative advantage relative to start-ups or smaller firms due to its deep financial pockets, and existing economies of scale as well as network effects (also in data analytics). In this way the incumbent can successfully prevent customer churn and, at the same time, deny start-ups a viable and stable customer base. Such 'kill zones' also seem to have an effect on the venture capital market, where start-ups that complement the incumbent's business model are more likely to receive venture capital than start-ups that challenge the incumbent (for a discussion see, e.g., Smith 2018, Rinehardt 2018 and Kamepalli et al. 2020). For the same reasons, there is also a growing consensus that data-intensive mergers should be reviewed more carefully and with adapted tools by competition authorities, (see, e.g., Bourreau and de Streel, 2020; Crémer et al. 2019; Motta and Peitz 2020; Furman et al. 2018; Scott-Morton et al. 2018).

In summary, it is inconclusive whether data portability would lead to more or less competition and innovation in established digital markets per se. It may, however, spur innovation in complementary and emerging digital markets, which we argue next.

4.2 Innovation at the service level vs. innovation at the analytics level

In Section 3.2.2 we have already discussed the positive feedback loop that provides an incumbent digital service provider with a competitive advantage in terms of data analytics capabilities. We now return to this issue from an innovation perspective. Data (volunteered and observed) is often accumulated as the results of *innovation at the service or product level*, which led consumers to use and thereby to contribute personal data. By contrast, inferred data is the result of *innovation at the data analytics level*. An important observation in this context is that, given the raw data, it does not necessarily require an innovation at the service level per se to achieve an innovation at the data analytics level.

However, as discussed previously, innovations with respect to inferred data (i.e., data analytics innovations) rest upon the input of raw data (observed and volunteered data), which typically can only be amassed if the firm also runs a successful service at the service or product level. This creates a virtuous innovation cycle for incumbents. Innovations at the analytics level facilitate innovation at the service level, which again spur innovations at the analytics level. While there are certainly

inherent efficiencies in this virtuous cycle, it may be viewed as problematic that innovation can, to a large degree, only occur 'in house', whereas truly innovative ideas often come from outsiders, frequently (business) users (see, e.g., van Hippel 2005). Indeed, innovation at the data analytics level may spur innovation at the service level in a completely different domain (Prüfer and Schottmüller, 2017). For example, Google Flu Trends[4] exemplified that search data cannot just be used to improve the search engine's results, but also to predict the spread of the flu. But it has also been demonstrated that there was significant scope for improvement over Google's algorithm (see, e.g., Lampos et al. 2015).

Similarly, an innovation at the service level may not get off the ground, if it is not fed with sufficient raw data to begin with. For example, collaborative-filtering based recommender systems suffer from a well-known 'cold-start problem' (see, e.g., Bobadilla et al. 2012). That is, in order to provide good results, the recommender system needs to be fed with sufficient user data (observed and volunteered data) in order to be able to find similarities between users from which recommendations can then be derived. For example, suppose it were an innovation at the service level to offer customers personalised recommendations for clothing and styling. If the idea is found to be intriguing enough by potential customers, it would – at least at the beginning – not be required to be very innovative at the analytics level, because collaborative-filtering algorithms for such a purpose would be readily available. The main challenge would be to overcome the cold-start problem, so that if new customers try the service for the first time, it would already offer useful recommendations.

Thus, there is reason to believe that innovation activities would be significantly increased, if it were possible that innovation at service level and innovation at the analytics level could occur independently, i.e., in different organisations. Thanks to the non-rivalry of data, this would not mean that the current data controller loses access to the data, and thus, can continue to be innovative both at the service *and* the analytics level, taking advantage of the virtuous feedback cycle.

4.3 Lack of empirical studies on data portability and innovation

While it is without doubt that we have seen an unprecedent wave of innovations in digital markets, the above arguments provide some reasoning that the level of innovation could be even higher, if data portability were more prevalent. To be clear, we are not aware of conclusive empirical evidence

[4] https://www.google.org/flutrends/about/

that has tested this hypothesis. In fact, while there is a substantial legal literature on data portability and some theoretical work (see Section 3), we are not aware of any empirical studies on how data portability specifically has altered competition or innovation incentives in digital markets. It is probably also difficult to establish a conclusive cause-and-effect relationship at all, as data portability usually comes in package with other privacy rights, and because in the dynamic environment of digital markets it is very difficult to establish the counterfactual for innovation.

There is some tentative evidence, however, in the case of Open Banking, which is probably one of the most important natural experiments in this context. Although there was competition between banks, the emergence of new financial services (fin techs) has spurred following the availability of API-based common interfaces that made continuous data portability possible (Open Banking, 2019). This seems to suggest that data portability has at least facilitated innovation activities in this sector.

5 The economics of Personal Management Information Systems

Personal Information Management Systems (PIMS) come in a variety of shapes, but their central premise is to empower users to regain control over their personal data. In this sense, PIMS are a catch-all term that represent any technical tool that helps to address the imbalance of power and transparency about data use between the individual, and the firm collecting its personal data. However, the core vision of PIMS is to provide users with a central dashboard, where they can manage their privacy rights. In particular, this means granting and revoking their consent for data processing — at a fine-grained level – with any given data controller, and exercising their legal rights, especially the right to data portability (Art. 20 GDPR) and the right to erasure (Art. 17 GDPR). Given their possibly central role in the context of data portability, we now discuss PIMS from an economic perspective. Particularly, we will focus on the questions whether and under which conditions PIMS may indeed be economically sustainable.

5.1 Key functionalities of PIMS

In policy and technical circles PIMS are often regarded as the silver bullet for empowering Internet users and for building a fair and transparent data economy. DG Connect published a report on PIMS already in 2016 (European Commission, 2016a), and the idea is still prominently discussed in the European Commissions' recently adopted Data Strategy (European Commission, 2020). The idea of

PIMS is much older, however, and dates back to the mid 90s when Laudon (1996) envisioned the creation of a national information market, where data subjects can deposit their information in bank-like institutions and are compensated for the use of their data.

However, to this date, the market for PIMS is highly dynamic, and many operators struggle to find a sustainable business model or steady customer base. Nevertheless, a myriad of different PIMS exist, and a comprehensive market overview would not only be beyond the scope of this paper, but also be likely outdated the date that it is published. We therefore refer to the website of the MyData movement, which originated in Finnish policy circles in 2014, and is collecting a case library of various PIMS initiatives on their website.[5] More specifically, according to a review of proto-typical PIMS by MyData (Langford et al., 2020), the key functionalities of PIMS can include:

- *Identity management*: Authentication at various services
- *Permission management:* Overview of data transactions and connections, including management of legal rights and consent
- *Service management*: Linking various data sources
- *Value exchange*: Accounting and capturing the value of data, including remuneration (personal data broker)
- *Data model management*: Managing semantic conversions (schemas) from one data model to another
- *Personal data transfers*: Implementing interfaces (APIs) for standardised and secure data exchange between various data sources and data recipients
- *Personal data storage*: Storing data from various sources, including data generated directly at the PIMS.
- *Governance support*: Ensuring compliance with legal frameworks
- *Logging and accountability*: Keeping historic logs of any data access and exchange facilitated by the PIMS

However, it is not always useful for PIMS to offer all of these functionalities. For example, as we will highlight below, whether or not PIMS should engage in value exchange (which we discuss later as *personal data brokers*) is debatable. It is also noteworthy that in the context of the digital economy,

[5] Available at https://mydata.org/cases

some of the key functionalities of PIMS are currently offered by large digital platforms directly. For example, the Data Transfer Project[6], which is backed by some of the largest digital platform providers, is a PIMS focused on personal data transfers and data model management. But possibly more importantly, large online platforms also offer online identity management solutions, i.e., registration and authentication of a user at various online services. For example, this is currently offered like Google, Facebook, Amazon, Microsoft, LinkedIn or Twitter. Thereof "Sign in with Google" and "Sign in with Facebook" are probably the most well-known.[7]

This begs the question whether, in the context of the digital economy, PIMS stand a chance to operate independently of the big tech firms, as neutral stand-alone brokers that can truly empower users to exercise control over their personal data. We explore this issue from an economic perspective in more detail below.

5.2 Lack of (de-facto) standards and APIs

The central premise of PIMS for users is that they offer a centralised dashboard that seamlessly integrates with the various services that they are using, offering key functionalities such as identity management, permission management and data transfers. This requires a common set of de-facto standards and high-performance APIs through which a PIMS would be able to access the various services and users' data. However, to date, such common standards are lacking. Instead, data integration is rather done through individual solutions, customised for each service, either using existing APIs (with rate and other access limitations) or through web scraping. Some view this as a central role of PIMS, because in this way PIMS enable some limited portability in an otherwise incompatible and non-standardised data ecosystem. However, we view this with some scepticism, because this approach is not scalable to a large set of services, and access hinges on the goodwill of the data provider, who may at any time make changes to the data access or undermine it. At the same time, without a significant user base, any given PIMS does not have sufficient leverage to set a standard on its own. By contrast, large digital platforms, such as Google or Facebook, have successfully leveraged their vast user base to induce many independent service providers to implement their standards, such as their single-sign-on solutions.

---

[6] See https://datatransferproject.dev
[7] See, for example, https://www.avg.com/en/signal/is-it-safe-to-log-in-with-facebook-or-google

In theory, widespread availability of APIs or a common export standard would alleviate this problem, because then network effects do not matter anymore. As long as all entities (PIMS and data controllers) can communicate thanks to common standards and interfaces, even a PIMS with only a few customers would be able to offer its consumers a comprehensive service. Several PIMS could even co-exist and thus PIMS could even compete for customers, as switching would be easy, due to the common standards, APIs, and the right to port volunteered data.[8] In this context, one must differentiate, however, at least between the i) standards for managing consent, ii) standards for authenticating the user, and iii) standards for actually transmitting and possibly storing personal data in order to enable key functionalities of PIMS.

While OAuth (Hardt, 2012) seems to be the de-facto standard for authentication, which is also used by "Login with Google" or "Login with Facebook" as well as in the Data Transfer Project, there are many implementation details that yet need to be considered (Data Transfer Project, 2018). Even with a common standard, centralised controlled could be retained, e.g., through the centralised control of crucial resources (such as token management in the context of OAuth) and rate management of APIs. Yet, in the other two 'standardisation domains' implementation approaches are even more isolated, and there is currently an ongoing development and debate how to design such standards. Recently, solutions based on blockchain designs have surged (see, e.g., Zyskind and Nathan 2015 as well as several industry initiatives[9]), because these promise a decentralised framework that could do without a centralised control and oversight. It yet remains to be seen, however, whether these solutions are practical and scalable.

5.3 Lack of sustainable business models

There also seems to be a lack of a sustainable business models for PIMS. Indeed, if we look beyond the need for standards and API access to connect a user's various data sources in a centralised PIMS, the question arises how the business model of a privately-financed 'neutral' data broker can ever be sustainable. In principle, there are three potential sources of revenues for a purely privately-financed PIMS, i) data markets, ii) data controllers and iii) consumers. If all of these turn out to be not

[8] We note that consent notifications would qualify as volunteered data.
[9] For example, by Microsoft (https://qz.com/989761/microsoft-msft-thinks-blockchain-tech-could-solve-one-of-the-internets-toughest-problems-digital-identities/) and Orbiter (http://www.orbiter.de/english/)

sustainable, there may also be a role for public subsidies. We discuss each of these possibilities in turn.

5.3.1 Generating revenue from data-driven services or on data markets

By data markets we mean any market where (access to) data or insights derived from data can be monetised. In particular, this can be advertising markets, the market for customer analytics services, and the market for data intermediaries (selling access to raw data). PIMS would then generate revenues in much the same way as the original data controllers (such as Google or Facebook) from which the data was transferred to the PIMS. Why would consumers then want to transfer their data to the PIMS at all? We see three possible reasons that warrant a discussion:

First, in this way consumers could exert some competitive pressure on data-rich platforms. In theory, PIMS could even have better data on its customers than any given data controller, precisely because PIMS have the possibility to aggregate data from various data controllers. That is, data sets might have greater 'depth' (i.e., more detailed user profiles). In practice this is not very likely, however. At least not compared to large online platforms which have the ability to track consumers' activity across multiple websites and services. In reverse, PIMS can only sell data from consumers that use the PIMS, and thus, data sets have less 'breadth' (i.e., less distinct user profiles). Even if the PIMS would have the same ability to generate insights from data and to offer data-intensive services, the extent to which PIMS can indeed exert competitive pressure and be a successful actor on the data markets is not clear. In this context, it is important to recall our discussion on the potentially fierce competition that could come along with selling identical data sets (see Section 2.3.2).

Second, users would have more control over where and which data is sold. This could be an incentive to transfer data to the PIMS in its own right. However, this additional control and transparency arises only with respect to the additional data sales by the PIMS, and not with respect to those data sales done by the data controller from which the data was transferred. Thus, if privacy is of concern to users, they first create an additional problem (selling more of their data) which they can then partially fix. This does not seem to be a very convincing incentive for consumers to transfer data. This may change, however, if, like in the California Consumer Protection Act (CCPA), consumers would additionally have the right to opt out of the sale of their personal data at the original data controller.

Precisely, CCPA (Cal. Civ. Code § 1798.135(a)(1))[10] states that a business that falls under the CCPA[11] shall

"Provide a clear and conspicuous link on the business's Internet homepage, titled "Do Not Sell My Personal Information," to an Internet Web page that enables a consumer, or a person authorised by the consumer, to opt-out of the sale of the consumer's personal information. A business shall not require a consumer to create an account in order to direct the business not to sell the consumer's personal information."

If consumers had the same right under European law, this would mean that a consumer could deny the original data controller to sell its data, and transfer it to a PIMS, who would then sell the data respecting the user's fine-granular control and consent options. This would indeed offer consumers more control over which data and to whom data is sold. PIMS could even compete with each other on the basis of finer control rights for the sale of data.

However, this would likely induce the original data controller to also offer consumers finer control rights with respect to how their data is sold – instead of just the full opt-out mandated by law. This, in turn, would give consumers less incentives to port their data to the PIMS in the first place. Consequently, from this view – and only if CCPA-like regulation would be adopted in Europe as well – PIMS could induce large online platforms to give users more control rights over how their data is used, because the market would become more contestable; but under this view, PIMS would probably never actually have a significant amount of customers, and would only serve as a competitive threat to achieve market contestability according to the contestable markets theory (Baumol, 1985). It is questionable whether this business model is sustainable, especially if setting up a PIMS involves significant fixed costs or venture capital, because PIMS would constantly be in a potential 'kill zone' in the following sense: Everything else being equal, consumers would find it easier to control their data directly at the original platform than to port it to a PIMS first. This gives the original platform a competitive advantage over a PIMS that would allow it to foreclose the PIMS from entry.

[10]See https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?lawCode=CIV&sectionNum=1798.135
[11] The CCPA applies to any business, including any for-profit entity that collects consumers' personal data, which does business in California, and satisfies at least one of the following thresholds: i) Has annual gross revenues in excess of USD 25 million; ii) Buys or sells the personal information of 50,000 or more consumers or households; or iii) Earns more than half of its annual revenue from selling consumers' personal information.

Nevertheless, the threat of entry by a PIMS remains, depending on shadow costs of entry, and disciplines the incumbent accordingly.

Third, and probably the most important incentive for consumers to transfer data to a PIMS under this revenue-generation scheme, is that the PIMS could pay consumers for their data. In other words, the PIMS would become a *Personal Data Broker (PDB)*, who sells personal data on behalf of the users, and offers users financial rewards in return (also called value exchange above). Consequently, PDBs are not just promising users more control over who they sell the data to, but foremost that users can financially participate from the commercialisation of their data. This is also the vision that was expressed already by Laudon (1996) and later by Larnier (2014), who also coined the term "data as labor" (Arrieta-Ibarra et al., 2018). Indeed, such PDB business models are currently being pursued in practice, e.g., by the joint venture between digi.me and UBDI (which stands for "Universal Basic Data Income")[12]. However, similar previous PDBs, such as Datacoup[13], have already failed and paid consumers only minimal rewards. According to Wikipedia[14] in the trial phase, Datacoup offered each user up to USD 5 per month, and in the beta phase up to USD 8 per month in return for access to user accounts of various social networks such as Facebook and LinkedIn, as well as to debit and credit card transactions. However, in November 2019 Datacoup announced its users that it is closing down, and had actually never sold any of their data up to this point. Instead, all payments had been made from the Datacoup treasury account. Other examples of PDBs are people.io (who seem to face similar issues[15] as Datacoup), Datum[16] (where data can be sold in return for cryptocurrency), ItsMyData[17] (which plans to pay consumers in the future, but does not do so yet[18]), and Wibson[19] (where users can earn tokens that can be redeemed in a marketplace; the market place has not been launched yet, however[20]). Even the large telecom operator Telefonica announced a PIMS with

[12] See https://www.marketplace.org/shows/marketplace-tech/an-app-that-pays-you-for-your-data-yes-actually/

[13] See https://www.datacoup.com
[14] See https://en.wikipedia.org/wiki/Datacoup
[15] See https://uk.trustpilot.com/review/people.io
[16] See https://www.datum.org
[17] See https://itsmydata.de/?lang=en
[18] See https://www.faz.net/aktuell/wirtschaft/digitec/start-up-it-s-my-data-moechte-die-demokratisierung-der-daten-16328619.html
[19] See https://wibson.org
[20] See https://medium.com/wibson/wibson-update-01-03-2020-352e9a422438

PDB in 2017, which they intended to call 'Aura'[21], but this project has never taken off the ground either.

Thus, while there is an emerging offer of PIMS that promise consumers to redeem them for their data (in the future), none of them currently seem to have a sustainable business model. Rewards are either very low or not being paid out yet. This is also in line with the game-theoretical model by Haberer, Krämer and Schnurr (2019) who show that the incumbent platform will strategically react to the emergence of PDBs by adapting the quality of its online service. In cases where the PDB is a relatively weak competitor on the data market (i.e., the PDB is not very successful in monetizing user data on the data market), the PDB is either foreclosed by the incumbent, or will only be able to pay out a minimal reward. Overall consumer welfare will decrease in this range, because the incumbent platform reduces its quality in order to deter the PDB. Consumers benefit only if the PDB is a relatively strong competitor (i.e., is very successful in monetizing user data). In this case, the PDB pays users a positive and significant reward. However, in this case the platform will also start to charge users for access and not offer its service for 'free' anymore. In this way, the platform can appropriate some of the additional consumer surplus that was created by the PDB. This highlights that PDBs may well change the business model of incumbent platforms from a free (e.g., advertising based) to a paid (e.g., subscription based) business model.

Moreover, paying users for their data also gives rise to an ethical issue. Such practice would quickly reveal that the data of some users is more valuable than the data of others. Even worse, the 'valuable users' are likely to be the most economically advantaged anyway. One interesting feature of the current zero-price (ad funded) business models in the digital economy is that everyone can access the same services, irrespective of how valuable their own data actually is. PDBs could change that and indeed, some low value users might find they have to start paying for services that were previously 'free', whilst high value users get paid to use them.[22]

Relatedly, Bergemann, Bonatti and Gan (2020) as well as Acemoglu et al (2019) highlight the 'social dimension' of data, which reduces the value and monetary compensation for individual data points.

---

[21] See https://www.ft.com/content/3278e6dc-67af-11e7-9a66-93fb352ba1fe

[22] This, of course, does not mean that ad-funded services do not have other ethical or economic issues on their own. In particular, due to targeted advertising certain types of ads will only be shown to certain demographics, which may disadvantage specific groups disproportionally, e.g., in the context of ads for cheap consumer credit or political advertising.

Their argument is that data revealed by one individual also reveals information about other, similar individuals. This creates a data externality. When similar users have already revealed data to a data intermediary (a platform, or a PIMS), then the value of additional data by similar users is lower. This leads to an unravelling, whereby consumers with the lowest privacy preferences sell their data first, so that the data intermediary can acquire (statistical) information about users at relatively little costs. This social externality of data fundamentally undermines the idea that 'data ownership' of one sort or another actually empowers consumers to receive a 'fair' and significant remuneration for their personal data.

5.3.2 Generating revenue from data controllers

An alternative way to generate revenues for PIMS is to offer online service providers a convenient tool by which they can be compliant to the seemingly complex and evolving legal frameworks that have been established by GDPR, CCPA, and others yet to come. In this case, the PIMS serves as compliance service, which is to the benefit of the user (who can exercise his or her rights conveniently) and of the online service provider (who does not have to worry about compliance).

Such a business model is pursued, e.g., by Datawallet.[23] Interestingly, Datawallet initially started out with the idea of a PDB in the sense discussed above. However, the company recently shifted focus and now clearly advertises itself as a compliance tool for service providers. The revenue model rests exclusively on charging service providers, but not on charging consumers. Nor do they seek to make money by selling user data on their own.

It is unlikely, however, that this business model will attract the current data rich firms as customers. Large online platforms have sufficient scale to handle compliance with GDPR and CCPA on their own. Thus, the business model is clearly targeted at small and medium sised services and in this sense a welcomed addition to the data ecosystem. However, PIMS pursuing this business model will have little impact on the data ecosystem for personal data, because they do not exert competitive pressure on large data rich firms. This also means that this business model may well be sustainable, because it is unlikely that such PIMS are entering the 'kill zone'.

[23] See https://www.datawallet.com

5.3.3 Generating revenues from users

Some observers have noted (e.g., Section 4.3.3 of the Opinion of the German Data Ethics Commission 2020) that any business model that depends on generating revenues from profit maximizing data controllers is problematic per se. PIMS should act in the best interest of consumers, and not in the best interest of those that handle or monetise consumers' data. Therefore, business models that collect a flat subscription fee from users are preferred, because they do not rely on the type or amount of data handled by the PIMS. Again, the question is how sustainable such a business model would be. Especially if PIMS rely on a common set of standards, and therefore entry costs are relatively low, competition between PIMS that rely only on a flat subscription fee from users is likely to be fierce. At the same time PIMS should offer a secure and reliable architecture for controlling personal data, and should not see cost-cutting as their primary concern to stay in business. This tension may only be resolved by effectively limiting the number of PIMS available, e.g., through licensing.

5.3.4 No revenue generation

The preceding discussion highlighted that privately funded PIMS may either not be sustainable or not have a significant impact on the Internet data ecosystem. This may give rise for governmental intervention or PIMS which are not financed privately. If PIMS are indeed seen as a central element to empowering users, state subsidised or even state-run PIMS may in fact be the only option to address this market failure.

However, two potential caveats of state-run PIMS are worth mentioning here. First, the state is often a bad investor and innovator compared to private firms. This seems especially problematic in a highly dynamic and complex environment like the data economy. Second, it is not clear – from the perspective of the users – that the state is the better controller of personal data. In some jurisdictions, consumers may have larger distrust in the government handling their data than a private firm. Although there may be technical solutions to ensure that data indeed remains private, and cannot be intercepted by the state (e.g., through cryptographic means or decentralisation such as through distributed ledger/blockchain solutions), it is not clear whether this is indeed a convincing argument for non-experts. Moreover, in some jurisdictions, such as the US, consumers have heightened privacy rights vis-à-vis the state compared to their privacy rights vis-à-vis private firms.

In the European Union, on the contrary, the state generally has larger basis of authorisation for processing citizens' personal data than private firms, although GDPR applies equally to both.

A final option may be to rely on open-source, not-for-profit solutions for PIMS. It is not unlikely that such solutions may emerge, particularly when there are agreed-on standards on which such solutions can be built. Ongoing projects, such as the Data Transfer Project are indeed examples for such open-source not-for-profit solutions. However, it should be noted that the Data Transfer Project is still in its infancy and only very little progress has been made since its inception in 2018. This is especially noteworthy, since some of the largest tech firms are backing this project. In order to put the development of the project in perspective, it is informative to compare it to other open-source projects that Google and Facebook support and in which they truly have a vested interest. A common metric to assess the size and activity in an open source-project is to count its lines of code, and its number of forks (i.e., the number of spin-off projects). At the time of writing, for the Data Transfer Project this lies in the range of 44.000 lines of code and hundreds of forks. By contrast, the open-source machine learning framework of Google, 'TensorFlow', is in the range of 2,5 million lines of code and 80.000 forks; and also Facebook's machine-learning framework 'Pytorch' has about 1 million lines of code and 10.000 forks. Finally, it is also noteworthy that Google is by far the most active contributor to the Data Transfer Project and accounts for about 80% of the total code. By contrast, Facebook, who has been very vocal on the promotion of data portability (Zuckerberg, 2019) accounts for only about 3% of the total code. This shows that even when pursuing the avenue of open-source projects, policymakers may need to take a more active role in facilitating the emergence and use of such PIMS, for example by setting common standards or by reducing information asymmetries though audits. We will return to this point in our policy recommendations.

6 Recommendations for increasing the effectiveness of personal data portability in the platform economy

The right to data portability under GDPR has been in effect for just about two years, but to date empirical and theoretical research on its economic consequences is scant. We have identified several issues throughout our discussion of the economic effects of the right to data portability in the context of the digital platform economy, which we summarise next. These issues give rise to a number of

policy recommendations on how data portability can be made more effective in this domain, which we discuss thereafter.

6.1 The issues

First, we highlighted that the collection of personal data is highly concentrated in the digital economy. The issue arises primarily with respect to observed data (tracking data, clickstream data, behavioural data) and to a lesser extent with volunteered data. Volunteered data also tends to be more static, whereas observed data has a more dynamic character, i.e., it is generated at a much higher rate. It is therefore primarily the access to observed data, which is seen as problematic under the current legal regime. We have advocated that observed data should be included in data portability requests. However, the static and infrequent nature of a data portability request often diminishes the usefulness of observed data for other applications. Here, a more dynamic and continuous data portability would be desirable to overcome this issue.

Second, we have also argued that widespread data portability, including both volunteered and observed data, is likely to render digital markets more competitive and innovative. While there is a lack of empirical studies to back or refute this claim, we have argued that freeing personal data from organisational silos would enable more decentralised innovation, which could also occur more independently at the service and the analytics level. We have also argued that, due to inherent concentration in the collection of observed data, it is desirable to have competition rather at the level of inferred data and analytics, but not in the collection of data. Taken together, this provides a strong rationale to facilitate data portability of `raw' user input data (i.e., both volunteered and observed), but not derived and inferred data, as much as possible. This will also likely require to educate consumers on their rights, to make the data available to them transparent, and to derive technical solutions (through PIMS or other means) so that data portability is just a click away.

Third, there are numerous technical difficulties that arise from different standards and data formats that may be used following a data portability request. In particular, the sending provider must not adhere to a certain standard and can change it at any given point in time. These uncertainties regarding standards and their perseverance can make it very costly for the new provider to offer an interface to import data. In return, this means that more stringent and common standards for data portability are a key to ensuring that data is more widely imported and used. The provisions in GDPR, which merely call for a "structured, commonly used and machine-readable format" are not enough.

If the same type of data (e.g., photos, videos, search logs) would be made available in the same format, irrespective of the provider, then it would be more feasible to develop and provide respective import adapters. A more widespread availability of such adaptors and re-usability of ported data would also raise awareness among users and encourage them to port their data. The transfer could further be facilitated by PIMS, who could perform schema mappings between various services.

Fourth, given the novelty of the right to data portability, firms also raise legal concerns that might arise when including data in data portability requests and when accepting data from other providers. This includes potential conflicts of rights, especially regarding the porting of data provided by the data subject on other data subjects (e.g., address books, or pictures in which other people are tagged). But legal concerns also arise with respect to liability issues, such as who is responsible if data is lost or modified in the transfer process. A recent White Paper on Data Portability by Facebook (Egan, 2019) summarises these legal concerns well. Some of these concerns may be addressed with the current legal rules. However, in order to encourage that more is included under the scope of data portability and that firms are more willing to import data, especially in the context of the digital economy, more legal certainty and guidance would be welcomed. Moreover, there may be a role for a regulatory testbed, where innovative start-ups accepting ported data, could work more closely together with the privacy-regulator in order to develop legally sound and economically viable solutions.

Fifth, we have highlighted that, from a technical perspective, PIMS are an important and welcomed addition to the data ecosystem because they can drastically reduce the complexity of data portability and consent management for users. However, the existing offers are still in its infancy and we have also raised doubts whether, from an economic perspective, PIMS may find a sustainable business model—especially if they are indeed acting as a neutral data broker. A minimum requirement to make PIMS feasible is to develop common standards and APIs through which PIMS can interact with the various services in a standardised and immediate way.

Sixth, to date there is limited evidence that data portability is widely used. Rather, we think that the root of the problem lies in the evident chicken-and-egg problem. Not at least for the reasons given above, currently very few providers do indeed accept ported data from users. If data is imported, it is often not done via the data set that a user has exported following a data portability request, but rather through existing APIs or other workarounds. In reverse, this means there is a lack of use cases for consumers to exercise their right to data portability. We believe that more continuous and

standardised data portability is key to overcoming this chicken-and-egg problem. Moreover, the experience from telecom markets (number portability) shows that portability became widely adopted when the consumer merely needs to give consent, but the (technical) details of exchange are deliberated by the sending and receiving data controllers directly according to some standardised process. The experience from other industries, foremost the Open Banking Order in the UK, highlights that third-parties often do see a value in importing data, and that data importing becomes more likely when standards are in place that allow for a continuous importing of data. In the case of Open Banking, after a slow start, there has been a continuous increase in both the number of third-parties accessing the available APIs as well as in the number of API calls being made.[24]

Taken together, we therefore see scope for improvement in three areas: (i) effective enforcement of the current legal framework, (ii) a new right for continuous, real-time data portability, tailored for the digital economy, and (iii) enabling PIMS through standards. We discuss each in turn.

6.2 Effective enforcement and clear scope of data portability under GDPR and DCD

A first set of recommendations entails effective enforcement and legal certainty on existing legal frameworks for data portability, particularly Article 20 GDPR. This is especially needed in the context of the digital platform economy, where the collection of personal data is ubiquitous, and often occurs in the form of observed rather than volunteered data, such as by tracking consumers' across several websites (see Section 2.3.1). This creates legal uncertainty not only for providers but also for consumers.

6.2.1 Legal certainty on the scope and trade-offs of the data portability right

Thus, a first priority for policymakers is to increase the legal certainty with regard to the scope and the limits of data portability under Article 20 GDPR. In the context of the digital economy, where data is always processed by automated means and every click is potentially recorded, the tensions between purpose limitation, data minimisation and data portability are particularly immanent. More guidance is needed on issues like:

- To what extent exactly is *observed data* to be included in a data portability request? As laid out under Section 4, a wider scope of data portability, including both volunteered and

[24] See https://www.openbanking.org.uk/providers/account-providers/api-performance/

observed data, is desirable to stimulate data-driven innovation outside the current silos and is covered by the GDPR according to the EDBP.

- In particular, does observed data include *location, tracking and clickstream data* (before being analysed or refined)? If so, how much context to such clickstream data should and needs to be made available so that data subjects can truly assess the information content of that data (e.g., exactly which content was consumed, exactly which ads were clicked on)? What are objective legal, economic or technical reasons not to make location, tracking and clickstream data available? For example, are concerns about data security and about a possible loss of reputation due to data leakage or misuse at the end of the receiving data controller admissible? When exactly is technically infeasibility admissible as a defence for data rich firm in the digital economy?
- Is there an obligation for data controllers to install measures and tools so that every data subject must make an explicit decision on whether they *consent or dissent* in case another data subject asks to port data that affects their data rights (e.g., if a photo is to be ported on which the data subject is tagged)? What about data subjects who do not have a contract with the data controller (but, e.g., a photo with their name tagged nevertheless exists with that data controller and is to be ported)?
- If some portable data affects data *rights of other data subjects* (and some of those data subjects have dissented to porting), does this mean that no data can be ported, or must the data controller offer to port at least the portion of the data that does not affect data rights of other subjects?

Legal clarity which is in line with the realities of the digital economy is needed so that Article 20 GDPR will be effective. We realise that at some point these questions can become so complex that a case-by-case analysis is necessary. In this case, it should be clear what are the main trade-offs and where firms and consumers can find legal guidance on the balancing of those trade-offs in a timely manner. In particular, in these cases, providers willing to facilitate data portability for consumers should be able to receive specific guidance by the privacy regulator in a cooperative approach. In this context, it is also worthwhile to discuss the use of *sandbox regulation*, which can provide a regulatory safe-harbour under which data portability can be developed further. Sandbox regulation under the auspices of a data protection authority was pioneered by the UK Information

Commissioner's Office, and is also successfully applied by the UK Financial Conduct Authority in the context of financial services.[25]

6.2.2 User-friendly transparency on data

A second priority is that there should be more transparency about the categories and extent of personal data that firms in the digital economy hold about a certain data subject. This information should be readily available to users already before a formal access request (Article 15(3) GDPR) or data portability request (Article 20 GDPR) is initiated. Data subjects already have these rights under Articles 12 and 15 GDPR, but currently there still seems to be, in some cases, a lack of transparency concerning the actual extent of data collection pertaining to each data subject (e.g., on the extent of tracking data). In our view, this information can be made more transparent and accessible to data subjects in the context of digital service providers, e.g., through the use an appropriate dashboard in the respective user's privacy settings. To be clear, several large online platforms, including Google and Facebook, already provide comprehensive dashboards.[26] However, other large online platforms, like Amazon, for example, do not. These dashboards could also be used to consent to data portability requests of other data subjects for individual data categories.

6.2.3 Effective monitoring and enforcement

A third priority is that there should be an effective monitoring and enforcement of the existing provisions on data portability under GDPR. This requires first that the scope and the limits of these provisions are clear in the context of the digital economy (see first priority) and that users are well aware about the data that is available about them and can be ported (see second priority). Then, there should be an effective monitoring and enforcement of the:

- timeliness in pursuing data portability requests relating to Article 12(3) GDPR,
- completeness of data (volunteered and observed data) in data sets created for portability,
- admissibility of technical feasibility constraints,
- admissibility of fees for data portability requests, particularly in the context of repeated requests relating to Article 12(5) GDPR.

---

[25] See https://www.fca.org.uk/firms/innovation/regulatory-sandbox for more details.
[26] See, e.g., Facebook (2020), and specifically https://www.facebook.com/your_information/ and https://www.facebook.com/off_facebook_activity/

6.3 The Need for Continuous, Real-Time Data Portability

Data portability under the scope of Article 20 GDPR, when clarified and enforced effectively as recommended in Section 6.2, is a welcome and necessary step to empower consumers to exercise their privacy rights. It should also facilitate switching from one digital service provider to another. However, one-off data portability according to Article 20 GDPR may not be sufficient to truly empower users in digital markets to foster competition and innovation. Often consumers want to try out a new service provider immediately, and that provider may be in need to cold start with the users' data in order to offer an immediately appealing service. But the GDPR does not give the consumers the right to immediate and very frequent access to their data. Consumers may have to wait up to a month or longer to receive the portable data from their current provider, and may face constraints regarding the frequency of these requests. Moreover, often consumers do not want to immediately switch to a new provider completely, but multi-home between providers first.[27] In this case a much more frequent porting of personal data would be desirable.

6.3.1 Objectives and legal tools

As we have noted in Section 2.3, some of the personal data generated by data rich firms in the digital economy will not be easy to replicate. The real advantage of many firms rich in personal data is that they can combine personal data from many different sources seamlessly and in real time in order to create detailed user profiles. A one-off data transfer with a delay of up to a month is not consistent with this reality. Article 20 GDPR was not intended for continuous (i.e., very frequent) data transfers which would empower users to port their data (particularly observed data) from one service to the various other services that they may be using in (near) real-time.

Moreover, many commentators agree that ex-post competition law is not the right instrument to address the data access and portability issues, especially if the purpose is to promote innovation (e.g., at the data analytics level) in general, and not to contest a specific market (Crémer et al., 2019). The legal barriers to obtain data access under competition law are typically very high (for good reasons), interventions take a long time and, most importantly, competition law is not well-suited to develop timely and effective remedies in the complex environments of digital markets (see inter alia, e.g., Crémer et al. 2019, Furman et al. 2019, Feasey and Krämer 2019). The advantage

[27] In this sense also Crémer et al. (2019, p.82).

of data portability is that it can offer personal identifiable consumer level data. But its disadvantage is that only a relatively small fraction of consumers will ever port data, such that the data is not representative. Thus, data access requests under competition law (or some other legal framework) will continue to play a role in the future.

In summary, we therefore argue in favour of a new proportionate rule that gives consumers the right to transfer their personal data (as under Article 20 GDPR) continuously and in real time from their existing digital service provider to another provider. This is what we refer to as 'continuous data portability'. This is not an entirely new policy proposal. It relates immediately to the 'Smart Data' initiative in the UK which, is initially focussed on regulated industries, beginning with the Open Banking, but also seeks to include digital markets in the future.[28] Recently, these efforts have been subsumed under the UK's National Data Strategy[29], which tackles the issue of 'data availability' more widely. Also in the specific context of online platforms, the UK's Competition and Markets Authority (2020) has stressed – in line with the recommendations of Furman et al. (2019) – that stronger data mobility interventions that allow consumers to share the data that platforms hold on them with other parties, would be a necessary and welcomed step for promoting both competition and consumer control over their data. Similar steps are being taking under the new Consumer Data Right (CDR) in Australia. The policy proposal also relates to the recently adopted European data strategy, who recognises that the "absence of technical tools and standards" makes the exercise of data portability burdensome (European Commission, 2020, p. 10). Indeed, even several of the largest tech firms recently expressed their efforts to give users more control over their data and privacy.[30] Facebook CEO Mark Zuckerberg explicitly urged governments for more regulation, and identified data portability as one of four areas where such action should be taken (Zuckerberg, 2019). The envisioned regulation on continuous data portability would be a step in this direction.

At this point, it is also important to reflect upon the potential risks that continuous data portability may bear. First, data portability can in itself create harms to the individual, because they may be induced (or even incentivised) to shift data about themselves from a well-controlled data processing

[28] See https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/808272/Smart-Data-Consultation.pdf
[29] See https://www.gov.uk/guidance/national-data-strategy
[30] See, e.g., https://eandt.theiet.org/content/articles/2020/01/google-ceo-backs-gdpr-says-privacy-should-not-be-a-luxury/

environment to a higher risk environment. For example, an aggressive start-up could pay customers to transfer data to them at the expense of appropriate security controls. To be clear, this concern applies already to the right to data portability as it stands now under Article 20, but it is likely to be amplified under continuous data portability. Thus, it may also be also worthwhile to consider a vetting procedure for firms that seek to import customer data continuously, as is the case in Open Banking, or at least that importing firms are registered, as is required by PSD2. In any case, customers must be able to control on a fine-grained level which data they seek to transfer, and firms should not be allowed to directly or indirectly influence a customer's consent based on financial incentives (see also 6.3.2). Second, continuous data portability also amplifies the legal uncertainties and risks raised in Section 6.2. For example, if liability and redress rights are unclear, customers are less likely to exercise their rights. Thus, the legal issues raised in Section 6.2. need to be thoroughly addressed first. Finally, there is an economic risk that the need to implement continuous data transfer would pose an undue burden, especially for start-ups and smaller firms. Therefore, and in accordance with the proportionality principle, the obligation to implement and enable continuous data portability should only be applicable when its benefits are likely to exceed its costs.

6.3.2 Guidelines for implementation

Lessons for the implementation of such an extended right to data portability can be drawn from the Free Flow of Data Regulation/FFDR (European Commission 2019d,e) and also from Open Banking in the UK (Ctrl-Shift, 2018). Like in the FFDR, as a first step, we propose a participatory, adaptative, and soft approach in the first phase. Thus, the regulation could require the establishment of codes of conduct (see, e.g. European Commission 2019e) and agreements on common standards and APIs, including performance criteria for the availability of these. Much in line with De la Mano and Padilla (2018), we suggest that the following points should be included in such guidelines in any case:

- Consumers must be able to give their *consent on a fine-granular level* regarding which data is to be transferred. All-or-nothing transfers are often not necessary, and would create more transaction costs, both technically (e.g., network load, space requirements) as well as economically (larger privacy concerns). They would also run counter the legal requirements of data minimisation under GDPR; firms shall not influence consent or disconsent by offering commercial incentives or disincentives.
- In line with Art. 20.2 GDPR, consumers should have a right to consent to a direct data transfer from the sending to the receiving firm. This means that also continuous, real-time

data portability should be possible without any additional infrastructure at the consumer end in order to reduce the complexity of the data portability process for consumers. However, this does not preclude the possibility that users employ PIMS to store data or to facilitate this process.

- Relatedly, the *nature and scope of the data ported should be very clearly communicated* to consumers, in plain language; generally, the scope of portable data should be the same as under Art. 20 GDPR.
- The continuous, real-time data transfer needs to be as *secure* as the one-off data transfer under Art. 20 GDPR, minimizing risks for data leakage to parties not involved in the transfer, data modification or loss of data;
- Where possible *open standards* and protocols, which are free to use and transparent for developers, should be used for continuous data portability (cp. Furman et al., 2019, pp.71-74);
- APIs need to be available with a *high reliability and performance*. Like under the PSD2 (European Commission, 2015), APIs should have the same performance and reliability as the interfaces that consumers otherwise use to interact with the digital service provider.

Several options are possible for the policy process by which these guidelines are transformed into technical solutions, ranging from industry-led self- and co-regulation to a standardisation body with legal powers. We believe that the development of standards and codes of conduct should be industry-led through multi-stakeholder groups, as in the case of the FFDR. All parties involved should negotiate in good faith to achieve the best possible outcome in the interest of the consumer. Given the international nature of digital services and data standards, possibly, the development can be facilitated by independent international standard setting committees, such as the W3C. However, as such committees typically require unanimous decisions, it needs to be taken care that developments are not vetoed by single parties to protect their market power.[31]

Ideally, the development of standards and technical solutions can be built on existing projects such as the Data Transfer Project. Of course, the devil is in the detail and implementing this involves

[31] This has occurred, for example, recently where Google blocked a vote to give the W3C's privacy group more powers. See https://www.bloomberg.com/news/articles/2019-09-24/google-blocks-privacy-push-at-the-group-that-sets-web-standards

challenges, as the implementation of PSD2/Open Banking or cloud-based services like IaaS and SaaS have shown. Given the demonstration project of Data Transfer Project, there does not seem to be a compelling technical reason why this is not feasible in a wider context. It is also to be expected that, once standards are defined and APIs are available, there will be a significant effort from the open-source community to provide import and export adapters between various services. There should be a timely deadline after which the progress and implementation status is evaluated by the competent authority.

If no sufficient progress has been made by means in establishing standards and operational interfaces within a specified period of time, there may be a need for stronger governmental intervention or guidelines to ensure progress is made and neutrality of interests are warranted. For example, in Open Banking the major banks were required to constitute an independent trustee to develop standards. In the case of PSD2, relatively detailed technical provisions were adopted by the Commission on the basis of the participatory work done at the European Banking Authority. Similar case-by-case provisions are also done in Australian Consumer Data Right (CDR) initiative.[32]

The ultimate option is to enact a public standards organisation to achieve this end. For example, the Australian government has given a legal mandate the Data Standards Body to develop standards for data access and portability.[33] It works in close collaboration with the competition authority and the data protection authority. Also the UK's Competition and Markets Authority (2020, p. 435) states that greater cooperation between the data protection authority (ICO) and the competition authority (CMA) is required in order to truly advance data mobility interventions in the future. Given the various ongoing data portability initiatives in the UK, it will be interesting to observe whether, from January 1, 2021 onwards, the UK will use its new legislative freedom to diverge from the EU with regards to its data protection and data mobility regime. If so, this will surely provide an interesting natural experiment from which important lessons can be drawn in the future.

6.4 Enabling and Governing Personal Information Management Systems

With a larger and continuous flow of personal data, facilitated by a right to continuously port data from large digital service providers, the role of Personal Information Management Systems (PIMS)

[32] See https://www.accc.gov.au/focus-areas/consumer-data-right-cdr-0
[33] See https://consumerdatastandards.org.au

is likely to become very important in practice. In particular, we consider PIMS as an essential tool for reducing the complexity of data portability for consumers. By contrast, other functionalities of PIMS highlighted in Section 5, like personal data stores or personal data brokers, are considered of lesser importance. This means, PIMS should at least provide a centralised management of user's privacy settings and consented data flows; ideally aggregating relevant information across the various digital services that the consumer is using, and being able to change settings across several services as needed. In this sense, PIMS would provide a dashboard of dashboards for user's privacy settings.

The role of PIMS should not be underestimated. In order to achieve effective data protection and data portability it is essential that consumers are aware of their given consents and exercise their rights, particularly if this is the basis on which data is being shared between firms.[34] In order to facilitate this, a centralised consent management is seen as crucial, as otherwise recent empirical studies suggest that this may lead to a vertical integration of PIMS with large platforms (Marthews & Tucker, 2019), which would run contrary to the intention of the PIMS being a neutral broker.

In order to enable a centralised consent management, additional standards for consent management need to be agreed over and beyond those needed for data transfers. Here, the same guidelines and recommendations as for the standards development for data transfer (Section 6.3.2) should apply. In order to facilitate this process at an early stage, additional funding for research and development on secure, decentralised and scalable solutions for consent management (e.g., based on blockchain technology) could be made available (cp. European Commission 2016a, p. 16).

Second, as we have pointed out in Section 5, even if standards for data portability and consent management are developed and the policy recommendations under 6.2 and 6.3 are being pursued, PIMS may struggle to find a profitable and sustainable business model. Indeed, it is crucial that PIMS become and remain a neutral intermediary acting purely on the behalf of consumers. This also why it has been suggested, e.g., by the German Data Ethics Commission, that there should be regulatory guidelines on acceptable business models for PIMS, preferring, e.g., business models based on flat monthly fees for consumers. Again, we pointed out in Section 5 that we are doubtful that such a

[34] In the absence of effective consent dashboards, it has been proposed that data portability consents should only be valid for a certain period of time. For example, in the context of PDS2, consumers need to renew their consent every 90 days. While this practice clearly reminds the consumer of the consents that have been given, it also increases the complexity and transaction costs of data portability for the consumers further.

business model would be sustainable, or would have a sizeable impact. Moreover, it raises the question whether PIMS should not be available free of charge to consumers, because otherwise, there would be a monthly price tag on consumers' privacy management, which may not be in line with European ethics values.

However, we also expect that if such standards are in place, there will be considerable development in open source communities, providing decentralised non-profit solutions. Given the potentially sensitive nature of the data being handled through PIMS, there may still be a need for public oversight, such as through *privacy seals and certification* (cp. European Commission, 2016a, p.12).

To achieve critical mass for PIMS, a fruitful avenue may also be to build a user base on top of existing or developing identity management solutions. In particular, the European Commission is currently pushing national governments to offer an interoperable European identity management based on public national electronic identification (eIDs).[35] Moreover, during the European Council Meeting held on 10th March 2020, Heads of States and governments agreed to launch an initiative entitled European Digital Identity, "with the aim of developing an EU-wide digital identity that allows for a simple, trusted and secure public system for citizens to identify themselves in the digital space by 2027". This could also be a starting point to couple *identity management* with *consent management*, and to link the eIDAS regulation to the Digital Services Act, which is expected in about the same time frame.

7 Conclusions

Personal data portability in the form of Article 20 GDPR was an important first step in empowering users to take their data, volunteered or observed, to wherever and whoever they wish. In the same spirit a wider movement of data mobility was stimulated and triggered, also encompassing non-personal data, such as in the Digital Content Directive (European Commission, 2019a) or Free Flow of Data Directive (European Commission, 2018).

The specific impact of data portability, as it is implemented in the legal framework today, is not yet clear, since there is a lack of empirical studies on this topic. Nevertheless, it is evident that many data services do not yet offer import possibilities for ported data, and consequently, data portability

[35] See https://ec.europa.eu/digital-single-market/en/trust-services-and-eid

in the legal sense is not used widely by consumers in the digital economy. Many commentators have argued that this is not due to the fact that data cannot be used outside of the context where it has been created, but rather due to a lack of common standards and APIs to access the data in a convenient and timely manner. Moreover, legal issues have been raised with respect to liability and protection of the rights of others.

We argue in this article that all of these issues can be overcome, albeit with possibly considerable efforts. Regarding the legal issues, more specific guidance should be offered how data portability can be facilitated and which data is subjected to data portability in the digital economy without violating privacy rights. In particular, we advocate that observed data should clearly be included under the scope of data portability, and that a wide interpretation of observed data should be adopted, including clickstream and tracking data, if available. In order to make inherent trade-offs salient and in order to resolve them, an open and constructive dialogue between data-intensive firms in the digital economy and regulators is necessary. This could evolve around prototypical use cases for data types to be transferred (e.g., posts, videos, photos, search logs, clickstreams, geo-location, ad views). Eventually, data portability also likely requires explicit consent by consumers on data portability requests initiated by others that including their data; and in some cases more transparency on the personal data that is available for porting.

Further, we argue that continuous, real-time data portability, facilitated by common data standards and APIs is technically feasible, albeit challenging as well. In fact, many providers do have respective APIs in place already, either privately or publicly accessible, and often with considerable technical or commercial constraints. Also demonstration projects, like the Data Transfer Project, highlight that continuous data portability is technically feasible. We therefore consider it essential that, within the scope of digital platform markets, there should be a more widespread obligation to offer standardised APIs to enable consumers to continuously port their data. This echoes ongoing policies in the UK and Australia, and we believe that the European Commission, in its Digital Strategy, should follow suit. We believe that standardised APIs to enable continuous data portability are a necessary prerequisite to encourage more firms to import personal data, and to encourage more consumers to initiate such transfers. Eventually, this is likely to spur innovation and competition in digital markets, although it is not likely that existing market structures are disrupted. Indeed, such an obligation must always keep a proportionality principle in mind, in order not to be overly burdensome for small and emerging digital service providers.

We also believe that Personal Management Information Systems (PIMS) have a crucial role to play in a digital economy where data portability is adopted widely. In particular, PIMS should facilitate the complex consent management and offer a centralised dashboard where the various data flows can be controlled. Data portability must be as easy as possible for consumers in order not to overwhelm them with data portability choices that may ultimately prevent them from exercising their rights. This may well require to further educate and inform users about their rights in information campaigns alongside with the policy measures that we have proposed here.

Moreover, we are sceptical that PIMS can be economically self-sustained and can find a business model in which they can act as a truly neutral agent, acting purely on behalf of the consumer. We sketch two avenues in which PIMS could be developed. First, standards for consent management could be established in order to enable PIMS to access and control the privacy settings in various services. Combined with appropriate certification, this could drive open-source solutions. Second, policy makers could strive for a more active role, and couple the development of PIMS and consent management standards more closely to its ongoing efforts for a joint European identity management solution.

In closing, it is important to highlight that a comprehensive policy agenda in the context of digital markets should not be limited to the goal of making data portability more effective, although this was the focus of this article. Policy makers should also consider how data sharing can be facilitated, and possibly mandated, more generally in order to foster competition and innovation, and hence consumer choice, in the digital economy as a whole. In many applications, customer data is especially valuable for competition and innovation when it is both *broad*, i.e., representative for the entire customer base, and *deep*, i.e., contains detailed information about each customer (Krämer, Schnurr & Broughton Micova, 2020). Data portability allows to share deep customer data, as it is based on individual consent. However, only a fraction of the customers will opt-in to porting their data to any given third party, and therefore the shared customer data is generally not representative. Thus, policy makers should also consider complementary sharing instruments that enable new entrants or smaller competitors to receive access to broad customer data that is representative. Access to such complementary data assets can be done, e.g., in the form of aggregated, and sufficiently anonymised, customer data that is shared through APIs, by use of federated learning approaches,

or through the use of data trusts.[36] Recent advancements in privacy enhancing technologies have facilitated such sharing in a privacy preserving way. Nevertheless, the data assets that are shared in this way remain to be less deep, because sharing was not based on individual consent, and thus the data cannot be personally identifiable. Therefore, even in the presence of other forms of data sharing, effective data portability will continue to play an important role as a complementary means to share deep user data.

References

Abiteboul, S., André, B., & Kaplan, D. (2015). Managing your digital life. *Communications of the ACM*, *58*(5), 32–35. https://doi.org/10.1145/2670528

Acemoglu, D., Makhdoumi, A., Malekian, A., & Ozdaglar, A. (2019). *Too much data: Prices and inefficiencies in data markets* (NBER Working Paper No. 26296). National Bureau of Economic Research. https://economics.harvard.edu/files/economics/files/acemoglu_spring_2020.pdf

Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, P. (2005). Competition and innovation: An inverted-U relationship. *The Quarterly Journal of Economics*, *120*(2), 701-728. https://doi.org/10.1093/qje/120.2.701

Argenton, C., & Prüfer, J. (2012). Search engine competition with network externalities. *Journal of Competition Law and Economics*, *8*(1), 73-105. https://doi.org/10.1093/joclec/nhr018

Arrieta-Ibarra, I., Goff, L., Jiménez-Hernández, D., Lanier, J., & Weyl, E. G. (2018). Should we treat data as labor? Moving beyond "free". *AEA Papers and Proceedings*, *108*, 38-42. https://www.doi.org/10.1257/pandp.20181003

Barker, A. (2020, February 26). 'Cookie apocalypse' forces profound changes in online advertising. *Financial Times*. https://www.ft.com/content/169079b2-3ba1-11ea-b84f-a62c46f39bc2

Baumol, W. J. (1986). Contestable markets: An uprising in the theory of industry structure. In *Microtheory: Applications and origins.* MIT Press.

[36] For a comprehensive overview of the technical and economic implications of possible data sharing remedies, we refer to Krämer, Schnurr & Broughton Micova (2020).

Beggs, A., & Klemperer, P. (1992). Multi-period competition with switching costs. *Econometrica: Journal of the Econometric Society*, *60*(3), 651-666.

Bergemann, D., Bonatti, A., & Gan, T. (2020). *The economics of social data* (Cowles Foundation Discussion Paper No. 2203R). https://dx.doi.org/10.2139/ssrn.3548336

Bobadilla, J., Ortega, F., Hernando, A., & Bernal, J. (2012). A collaborative filtering approach to mitigate the new user cold start problem. *Knowledge-based systems*, *26*, 225-238. https://doi.org/10.1016/j.knosys.2011.07.021

Bourreau, M., & de Streel, A. (2020). *Big tech acquisitions: Competition & innovation effects and EU merger control*. Centre on Regulation in Europe (CERRE). https://www.cerre.eu/publications/big-tech-acquisitions-competition-and-innovation-effects-eu-merger-control

Bundeskartellamt (2019). Bundeskartellamt prohibits Facebook from combining user data from different sources. Press Release. Available at: https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2019/07_02_2019_Facebook.html?nn=3591568

Competition and Markets Authority (2020). Online platforms and digital advertising. Market study final report. Available at: https://assets.publishing.service.gov.uk/media/5efc57ed3a6f4023d242ed56/Final_report_1_July_2020_.pdf

Crémer, J., de Montjoye, Y. A., & Schweitzer, H. (2019). *Competition policy for the digital era*. European Commission. https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf

Ctrl-Shift. (2018). *Data mobility: The personal data portability growth opportunity for the UK economy*. UK Department for Digital, Culture, Media & Sport. https://www.ctrl-shift.co.uk/reports/DCMS_Ctrl-Shift_Data_mobility_report_full.pdf

Data Transfer Project (2018). Data Transfer Project Overview and Fundamentals. White Paper. Available at: https://datatransferproject.dev/dtp-overview.pdf

De la Mano, M., & Padilla, J. (2018). Big Tech Banking. *Journal of Competition Law & Economics, 14*(4), 494-526.

The Economist. (2018, June 2). *American tech giants are making life tough for startups*. https://www.economist.com/business/2018/06/02/american-tech-giants-are-making-life-tough-for-startups

Egan, E. (2019). *Charting a way forward: Data portability and privacy* [White Paper]. Facebook. https://about.fb.com/wp-content/uploads/2020/02/data-portability-privacy-white-paper.pdf

Englehardt, S., & Narayanan, A. (2016). Online tracking: A 1-million-site measurement and analysis. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 1388–1401. https://doi.org/10.1145/2976749.2978313

European Commission. (2015). Directive (2015/2366/EU) on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC. *Official Journal of the European Union*, *L337/35*, 35-127.

European Commission. (2016a). *An emerging offer of Personal Information Management Services: Current state of service offers and challenges*. https://ec.europa.eu/digital-single-market/en/news/emerging-offer-personal-information-management-services-current-state-service-offers-and

European Commission. (2016b). Regulation (2016/679/EU) on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*, *L119/1*, 1-88.

European Commission. (2017a). *Guidelines of Article 29 Data Protection Working Party on the right to data portability* (WP 242 rev.01). https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611233

European Commission. (2017b). *The new European interoperability framework*. https://ec.europa.eu/isa2/eif_en

European Commission. (2017c). Commission Delegated Regulation (2018/389/EU) supplementing Directive (EU) 2015/2366 of the European Parliament and of the Council with regard to regulatory technical standards for strong customer authentication and common and secure open standards of communication. *Official Journal of the European Union*, *L69/23*, 23-43.

European Commission. (2018). Regulation (2018/1807/EU) on a framework for the free flow of non-personal data in the European Union. *Official Journal of the European Union*, *L303/59*, 59-68.

European Commission. (2019a). Directive (2019/770/EU) on certain aspects concerning contracts for the supply of digital content and digital services. *Official Journal of the European Union*, *L136/1*, 1-27.

European Commission. (2019b). Antitrust: Commission fines Google €1.49 billion for abusive practices in online advertising. Press Release. Available at: https://ec.europa.eu/commission/presscorner/detail/en/IP_19_1770

European Commission. (2019c). Directive (2019/1024/EU) of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information. *Official Journal of the European Union*, *L172/56*, 56-83.

European Commission. (2019d, May 29). Guidance on the regulation on a framework for the free flow of non-personal data in the European Union. *COM (2019) 250*. https://ec.europa.eu/transparency/regdoc/rep/1/2019/EN/COM-2019-250-F1-EN-MAIN-PART-1.PDF

European Commission. (2019e, December 9). *Presentation of codes of conduct on cloud switching and data portability*. https://ec.europa.eu/digital-single-market/en/news/presentation-codes-conduct-cloud-switching-and-data-portability

European Commission. (2020, February 19). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A European strategy for data. *COM (2020) 66*. https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

Facebook. (2020). *Accessing & downloading your information.* https://www.facebook.com/help/1701730696756992

Farrell, J., & Shapiro, C. (1988). Dynamic competition with switching costs. *The RAND Journal of Economics*, *19*(1), 123-137.

Feasey, R., Krämer, J. (2019). *Implementing effective remedies for anti-competitive intermediation bias on vertically integrated platforms*. Centre on Regulation in Europe (CERRE). https://www.cerre.eu/sites/cerre/files/cerre_intermediationbiasremedies_report.pdf

Furman, J., Coyle, D., Fletcher, A., McAuley, D., & Marsden, P. (2019). *Unlocking digital competition: Report of the digital competition expert panel*. Government of the United Kingdom. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/un locking_digital_competition_furman_review_web.pdf

Gans, J. (2018). *Enhancing competition with data and identity portability.* The Hamilton Project. https://www.brookings.edu/wp-content/uploads/2018/06/ES_THP_20180611_Gans.pdf

German Data Ethics Commission. (2020). *Opinion of the Data Ethics Commission*. https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN_lang.pdf?__blob=publicationFile&v=3

Google. (2020a). *Download your data*. https://support.google.com/accounts/answer/3024190

Google. (2020b). *Google APIs Explorer*. https://developers.google.com/apis-explorer

Gu, Y., Madio, L., & Reggiani, C. (2019). *Data brokers co-opetition*. SSRN. https://dx.doi.org/10.2139/ssrn.3308384

Haberer, B., Krämer, J., & Schnurr, D. (2020). Standing on the shoulders of web giants: The economic effects of personal data brokers. SSRN. Available at http://dx.doi.org/10.2139/ssrn.3141946

Hardt, D. (2012). *The OAuth 2.0 Authorization Framework* (RFC No. 6749). Internet Engineering Task Force. https://tools.ietf.org/html/rfc6749

Hagiu, A., & Wright, J. (2020). *Data-enabled learning, network effects and competitive advantage* (Working Paper). http://andreihagiu.com/wp-content/uploads/2020/05/Data-enabled-learning-20200426-web.pdf

Ichihashi, S. (2019). *Non-competing data intermediaries*. SSRN. https://dx.doi.org/10.2139/ssrn.3310410

Junqué de Fortuny, E., Martens, D., & Provost, F. (2013). Predictive modeling with big data: Is bigger really better?. *Big Data*, *1*(4), 215–226. https://doi.org/10.1089/big.2013.0037

Kamepalli, S. K., Rajan, R. G., & Zingales, L. (2020). *Kill Zone* (CEPR Discussion Paper No. DP14709). https://ssrn.com/abstract=3594344

Klemperer, P. (1987a). Markets with consumer switching costs. *The Quarterly Journal of Economics*, *102*(2), 375-394. https://doi.org/10.2307/1885068

Klemperer, P. (1987b). The competitiveness of markets with switching costs. *The RAND Journal of Economics*, *18*(1), 138-150.

Krämer, J., Schnurr, D. & Broughton Micova, S. (2020). The role of data for digital markets contestability: case studies and data access remedies. Centre on Regulation in Europe (CERRE) Policy Report. Available at: https://cerre.eu/wp-content/uploads/2020/08/cerre-the_role_of_data_for_digital_markets_contestability_case_studies_and_data_access_remedies-september2020.pdf

Krämer, J., & Stüdlein, N. (2019). Data portability, data disclosure and data-induced switching costs: Some unintended consequences of the General Data Protection Regulation. *Economics Letters, 181*, 99-103.

Krämer, J., & Wohlfarth, M. (2018). Market power, regulatory convergence, and the role of data in digital markets. *Telecommunications Policy*, *42*(2), 154-171. https://doi.org/10.1016/j.telpol.2017.10.004

Lam, W. M. W., & Liu, X. (2020). Does data portability facilitate entry?. *International Journal of Industrial Organization*, *69*, Article 102564. https://doi.org/10.1016/j.ijindorg.2019.102564

Lanier, J. (2014). *Who owns the future?*. Simon and Schuster.

Lambrecht, A., & Tucker, C. E. (2015). *Can big data protect a firm from competition?*. SSRN. https://dx.doi.org/10.2139/ssrn.2705530

Lampos, V., Miller, A. C., Crossan, S., & Stefansen, C. (2015). Advances in nowcasting influenza-like illness rates using search query logs. *Scientific reports*, *5*, Article 12760. https://doi.org/10.1038/srep12760

Laudon, K. C. (1996). Markets and privacy. *Communications of the ACM*, *39*(9), 92-104. https://doi.org/10.1145/234215.234476

Lerner, A. V. (2014). *The role of 'Big Data' in online platform competition*. SSRN. https://dx.doi.org/10.2139/ssrn.2482780

Lewis, R. A., & Rao, J. M. (2015). The unfavorable economics of measuring the returns to advertising. *The Quarterly Journal of Economics*, *130*(4), 1941–1973. https://doi.org/10.1093/qje/qjv023

Li, X., Ling, C. X., & Wang, H. (2016). The convergence behavior of naive bayes on large sparse datasets. *ACM Transactions on Knowledge Discovery from Data*, *11*(1), 1–24. https://doi.org/10.1145/2948068

Macbeth, S. (2017). *Tracking the trackers: Analysing the global tracking landscape with GhostRank*. https://www.ghostery.com/wp-content/themes/ghostery/images/campaigns/tracker-study/Ghostery_Study_-_Tracking_the_Trackers.pdf

Martens, D., Provost, F., Clark, J., & Junqué de Fortuny, E. (2016). Mining massive fine-grained behavior data to improve predictive analytics. *MIS Quarterly*, *40*(4), 869-888. https://doi.org/10.25300/MISQ/2016/40.4.04

Marthews, A., & Tucker, C. (2019), *Privacy policy and competition*. Brookings. https://www.brookings.edu/wp-content/uploads/2019/12/ES-12.04.19-Marthews-Tucker.pdf

McLeod, J. (2020, February 7). Inside the kill zone: Big tech makes life miserable for some startups but others embrace its power. *Financial Post*. https://business.financialpost.com/technology/inside-the-kill-zone-big-tech-makes-life-miserable-for-some-startups-but-others-embrace-its-power

Motta, M., & Peitz, M. (2020). *Big tech mergers* (CEPR Discussion Paper No. 14353). Centre for Economic Policy Research (CEPR). https://cepr.org/content/free-dp-download-31-january-2020-competitive-effects-big-tech-mergers-and-implications

Langford, J., Poikola, A., Janssen, W., Lähteenoja, V., & Rikken, M. (2020). Understanding MyData Operators. Available at: https://mydata.org/wp-content/uploads/sites/5/2020/04/Understanding-Mydata-Operators-pages.pdf

OECD. (2019). *Enhancing access to and sharing of data: Reconciling risks and benefits for data re-use across societies*. OECD Publishing. https://doi.org/10.1787/276aaca8-en.

Open Banking (2019). Open Banking 2019 Review. Available at: https://www.openbanking.org.uk/wp-content/uploads/2019-Highlights.pdf

Prüfer, J. & Schottmüller, C. (2019). *Competing with Big Data* (TILEC Discussion Paper No. 2017-006). https://ssrn.com/abstract=2918726

Richardson, L., Amundsen, M., Amundsen, M., & Ruby, S. (2013). *RESTful Web APIs: Services for a changing world*. O'Reilly Media, Inc.

Rinehart, W. (2018, November 7). *Is there a kill zone in tech?.* Techliberation. https://techliberation.com/2018/11/07/is-there-a-kill-zone-in-tech/

Schaefer, M., Sapi, G., & Lorincz, S. (2018). *The effect of big data on recommendation quality: The example of internet search* (DIW Berlin Discussion Paper No. 1730). http://www.dice.hhu.de/fileadmin/redaktion/Fakultaeten/Wirtschaftswissenschaftliche_Fakultaet/DICE/Discussi on_Paper/284_Schaefer_Sapi_Lorincz.pdf

Schweitzer, H., Haucap, J., Kerber, W., & Welker, R. (2018). Modernisierung der Missbrauchsaufsicht für marktmächtige Unternehmen. Nomos Verlag. https://doi.org/10.5771/9783845296449

Scott Morton, F., Bouvier, P., Ezrachi, A., Jullien, B., Katz, R., Kimmelman, G., Melamed, A. D., & Morgenstern, J. (2019). *Stigler committee on digital platforms: Market structure and antitrust subcommittee report*. University of Chicago Booth School of Business. https://research.chicagobooth.edu/-/media/research/stigler/pdfs/digital-platforms---committee-report---stigler-center.pdf?la=en&hash=2D23583FF8BCC560B7FEF7A81E1F95C1DDC5225E

Smith, N. (2018, November 7). *Big tech sets up a 'kill zone' for industry upstarts*. Bloomberg. https://www.bloomberg.com/opinion/articles/2018-11-07/big-tech-sets-up-a-kill-zone-for-industry-upstarts

Suleymanova, I., & Wey, C. (2011). Bertrand competition in markets with network effects and switching costs. *The BE Journal of Economic Analysis & Policy, 11*(1). DOI: https://doi.org/10.2202/1935-1682.2359

Tombal, T. (2018). Les droits de la personne concernée dans le RGPD. In *Le règlement général sur la protection des données (RGPD/GDPR): analyse approfondie*, (44), 407-557. Larcier. http://www.crid.be/pdf/public/8347.pdf

Tucker, C. (2019). Digital data, platforms and the usual [antitrust] suspects: Network effects, switching costs, essential facility. *Review of Industrial Organization*, *54*(4), 683–694. https://doi.org/10.1007/s11151-019-09693-7

Van Hippel, E. (2005). *Democratizing innovation*. MIT Press. https://ssrn.com/abstract=712763

Wohlfarth, M. (2019). Data portability on the internet: An economic analysis. *Business & Information Systems Engineering*, *61*(5), 551–574. https://doi.org/10.1007/s12599-019-00580-9

Wright, A., Andrews, H., & Hutton, B. (2019). *JSON Schema: A media type for describing JSON documents* (Internet Draft). Internet Engineering Task Force. https://tools.ietf.org/html/draft-handrews-json-schema-02

Zingales, L., & Rolnik, G. (2017, June 30). A way to own your social-media data. *The New York Times.* https://www.nytimes.com/2017/06/30/opinion/social-data-google-facebook-europe.html

Zuckerberg, M. (2019). The Internet needs new rules. Let's start in these four areas. Opinion by Mark Zuckerberg. Washington Post. Available at: https://www.washingtonpost.com/opinions/mark-zuckerberg-the-internet-needs-new-rules-lets-start-in-these-four-areas/2019/03/29/9e6f0504-521a-11e9-a3f7-78b7525a8d5f_story.html

Zyskind, G., Nathan, O., & Pentland, A. (2015). Decentralizing privacy: Using blockchain to protect personal data. *2015 IEEE Security and Privacy Workshops*, 180-184. https://doi.org/10.1109/SPW.2015.27