CPB Netherlands Bureau for
Economic Policy Analysis

Policy Options for the Data Economy - a Literature Review

Data play an important role in our economy. Our digital lives are increasingly intertwined. For example, a decision to share data often impacts privacy or user experience of others. Yet the decision-making power for sharing data often lies with individual consumers or organizations. How then to make the most out of the data economy?

In this background document, we review the economics literature on data and the data economy. We conclude that due to several market failures, collective action is a key ingredient of successful data policies.

CPB Background Document

Ramy El-Dardiry, Milena Dinkova, Bastiaan Overvest

May 2021

1 Introduction

This report serves as a background document to the CPB Policy Brief 'Brave new data' (2021). In this document we discuss the literature on the economics of data and potential policy options in more depth.

The rising importance of data in our economy and society has prompted more research into these topics in recent years. Data and the associated digitalization of our society present both tremendous opportunities and challenges. On the one hand, digitalization might introduce and sustain a new period of economic growth and help overcome societal challenges. For example, by enabling personalized education or preventive medicine. On the other hand, there are uncertainties about the future of privacy and our democracy, and worries about the power of a handful technological firms. New research provides a better understanding of how to think about these opportunities and challenges.

Given the wide and profound impact of digitalization, it is perhaps not surprising that research on data involves many disciplines. Economists have recently begun to better understand how data function as a factor of production. Legal scholars have leveraged concepts from economics to study externalities associated with data and privacy. Philosophers, sociologists and political scientists have been concerned with the new power balance that emerges from the digital era. Meanwhile, computer scientists are constantly inventing new ways to better safeguard privacy or to enable exchanges of data.

Furthermore, insights are generated in many places – e.g. universities, government agencies, think tanks, consulting firms – and disseminated via different means such as peer reviewed articles, blog posts and white papers. In this background document we review a great variety of sources, both from inside and outside academia. We focus on the economics literature, but frequently sidestep to other disciplines. We do not aim nor claim to give a complete overview. Rather we intend to give sufficient background material that supports the main conclusions of the Policy Brief.

The structure of the document follows the structure of the Policy Brief. First, we define and characterize what is meant with data in chapter 2. Here, we also review the economic properties of data such as non-rivalry and low replication costs, and estimates for the value of data in our economy. In chapter 3, we discuss the literature on market failures in the data economy. In particular, we study the literature on externalities, public goods, market power, information asymmetries and behavioral biases. In the final chapter, our foremost aim is to provide a solid basis for the policy options in the policy brief. To do so, we start by briefly reviewing ways to improve the current informed consent model. We also discuss recent proposals by the European Commission for new regulation. Then, to better understand how collective action might take place in the data economy we look into concepts that were developed in the literature on common pool resources. The economic rationale for data sharing and the different ways data sharing can materialize are reviewed. We conclude with a brief overview of policy options that put restrictions on certain parts of the data value chain.

2 Data characteristics

Data come in many different shapes and are used in a variety of ways. Understanding these differences is important for designing policies that balance opportunities and challenges. For example, using anonymized income statements for an academic paper on the financial performance of small and medium sized enterprises touches upon different issues than using someone's social media profile to target advertising. At the same time, some economic properties of (digital) data are independent of the data type. In this chapter, we first aim to get a better understanding of data by defining data, categorizing differences and identifying common denominators. Then, we study the data economy in more detail and discuss the value of data.

2.1 Definitions and categorization

In this section we discuss several papers from the economics literature that present either a classification scheme for data or provide definitions. A rich picture emerges that warns us for oversimplification when analyzing economic bottlenecks in the data economy or designing policies.

In an overview paper on the economics of data, Carrière-Swallow and Haksar (2019) define data as a “factual representation of a characteristic, action, or natural occurrence” (p.7). They make a distinction between qualitative and quantitative data and the way data is stored (digital versus analog). Hilbert and López (2011) show how data have become increasingly digitized during the last decades. Data are now predominantly stored digitally[1].

Data differ from ideas. Both are forms of information, but they serve different purposes. According to Jones and Tonetti (2020), ‘an idea is a production function whereas data is a factor of production.’ (p.2821) Concretely this means that ideas are pieces of information that provide instructions on how to create output from a certain set of inputs (Romer, 1990). Data on the other hand are used in the production process, either to create products or services or to create new ideas.

Several classification schemes for data have emerged in the literature (see Wdowin and Diepeveen (2020) for a more extensive overview). Crémer et al. (2019) make a distinction between personal and non-personal data, and classify data as volunteered, observed or inferred based on the channel through which the data have been acquired. Furthermore, they propose to distinguish between four categories of use cases. Applications and analyses can use individual-level data, bundled individual-level data, aggregate-level data or contextual data. Individual-level data refers to data from a specific user or machine. When individual-level data are combined, e.g. to come up with movie or music recommendations, they use the term bundled individual-level data. Without additional information, it is not possible to trace aggregate data back to the individual level. Examples include frequency tables showing the distribution of digital skills levels of a population group or profit and loss statements. Contextual data are not derived from individual-level data. Typical examples are satellite data, mapping data or earthquake data.

Statistics Canada (2019) suggests to organize data according to what they are about or what they represent – for instance weather data, sports data or economic data. In a report on international data transfers, the Swedish National Board of Trade (2014) classifies data based on how they are used in the production process of companies. Examples include employment data, quality data and customer data.

[1] For some fascinating ancient ways to store data see e.g. this BBC news article about the world's first accountants.

2.2 Economic properties

In this section, we study the economic properties of data. First, we focus on nonrivalry and partial excludability. Second, we discuss the impact of data on economic costs and the marginal returns of data.

2.2.1 Nonrivalry

One of the most distinctive features of data is nonrivalry. An economic good is nonrival when it can be used by multiple consumers or firms at the same time, without diminishing its quantity or quality. Jones and Tonetti (2020) use an illustrative analogy with rival goods to explain what it means that at 'the technological level, data is infinitely usable' (p. 2819). Because of rivalry, workers typically need their own desk and computer and every single warehouse relies on its own collection of forklifts. If we would assume this capital to be nonrival however, then all workers could use all desks and computers at once and all warehouses would be able to use any forklift in the industry. This is the case with data. Due to non-rivalry, all data could theoretically be used by all firms at the same time which implies that economic gains remain untapped as long as this nonrivalry is not exploited. Carrière-Swallow and Haksar (2019) note that policies and private interests eventually determine whether data will be nonrival in practice.

Goldfarb and Tucker (2019) generalize the nonrivalry of data to products and services by comparing goods made of atoms and goods made of bits. Unlike goods made of atoms bits are nonrival. This is because the replication costs of digital information are almost zero – you can copy-paste software code but not a Ferrari.

2.2.2 Partial excludability

Some types of data are excludable, i.e. denying others access is not prohibitively costly. When data collectors exclude others, data takes on the features of a club good (see Buchanan (1965) for a formal definition). When others cannot be prevented from accessing data, data is non-excludable and can be regarded as a public good.

Coyle et al. (2020) provide a short overview of the excludability of different data types. For instance, administrative data (like tax returns or patient records) or planned data (like work schedules or budgets) are types of data where others can easily be excluded from. In contrast, environmental data, such as rainfall data or geospatial data, are accessible to anyone since everyone can collect their own data of publicly observable phenomena – although the private costs of measurement may be too high to actually do it. A common way to make data excludable is by putting data behind a paywall (often tied to account registration). Think of newspaper articles or datasets for researchers. Offline storage is probably the easiest way for limiting access – only breaking physically into the device or space where the data is stored can lift the lock.

Data collectors and data processors face different incentives when deciding the level of access to data. They can for example restrict access in order to secure their competitive advantage and maintain their current market position (Carrière-Swallow and Haksar, 2019). Privacy legislation could be another reason for an organization to exclude others from access.

2.2.3 Impact of data on economic costs

Goldfarb and Tucker (2019) describe how digitalization reduces five economic costs (search costs, replication costs, transportation costs, tracking costs and verification costs). These reductions are all connected to the properties of digitized data.

First, digitalization decreases search costs. Search engines for example have made it much easier to find relevant information, whether it concerns products, knowledge or data itself. Second, replication costs of digital products are close to zero. In other words marginal costs are negligible. Moreover, as we have seen earlier, reproduction of data does not impact others due to its nonrival nature. Although marginal costs almost vanish, rolling out successful digital products often requires significant upfront investment – e.g. to establish a large enough network or to build a solid data infrastructure. Third, data are associated with near zero transportation costs. Data can be transferred across the globe without much effort. As a consequence digital business models have increasingly become global and businesses are able to scale at a more rapid pace. Fourth, tracking costs are lowered: digital data make it easier to keep track of transactions, people and firms. Digitalization has therefore led to increasing levels of personalization. Examples include price discrimination and personalized advertisements. Such ads have the potential to facilitate matching of supply and demand. Fifth, lower tracking costs have enabled a reduction in verification costs. Digital products and services have made it easier to verify identities and create reputation systems. Digital platforms, such as Uber and AirBnB, leverage the reduction in verification extensively to build trust in their two-sided marketplaces.

2.2.4 Increasing and decreasing returns

In their book *Radical markets* (2018) Eric Posner and Glen Weyl discuss the marginal value of data in depth. The marginal value of an extra data point can either be decreasing or increasing with the number of data points already collected, depending on the context.

To understand how this works, first consider a standard statistics problem. Let's say for example that you are interested in determining average household savings. The uncertainty in mean household savings decreases with the number of data points collected, but the marginal decline becomes increasingly smaller as more data points are added. Thus, data lose their value over volume and variety. Moreover, there is always a level of uncertainty that suffices for the application at hand. Gathering more data once this uncertainty level is reached is inefficient.

Posner and Weyl explain how in the data economy, where machine learning algorithms play an increasingly important role, marginal values of data can be increasing. The underlying reason for this increasing returns of data is that different algorithms require different amounts of data. Typically, more data are needed the more complex a problem is. For a single learning problem, data again exhibit diminishing value of return, but collecting more data might now enable *new* problems to be solved causing a jump in the value of data collected. Whether data have an increasing or diminishing value of return is then determined by the value of the different problems. When most value resides with the most complex problem, it is likely that data have increasing value of return. In contrast, when most value resides with the simplest problems, data are likely to have diminishing value of return.

2.3 The value of data in the economy

2.3.1 Extracting value from data: the data value chain

Data are input to production processes. The data value chain describes how data contribute to production. In our policy brief, we split up the data value chain into three (see Figure 1). First, data need to be collected and stored. Second, data get analyzed and combined to create insights. Third, the insights translate into products and services.

Firms and institutes in the data economy either focus on a part of the chain or control the entire value chain for their business. Cloud services and big data consultants are examples of firms that specialize in offering products for a particular part of value chain. The activities of tech firms, e.g. Alphabet, Amazon and Apple, span the entire value chain. In the literature different versions of the value chain appear. Differences originate from the number of chains or slightly different terminology.

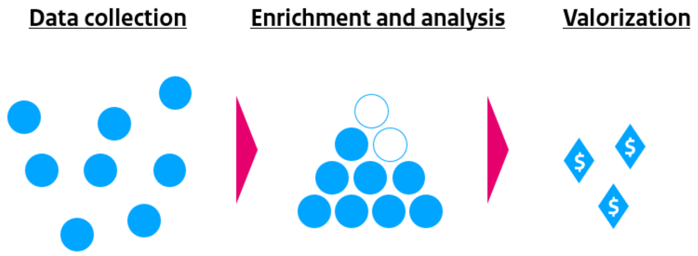


Figure 1: Data value chain

Often, economic agents who play a role in the value chain are referred to as data subjects, data collectors and data processors (Carrière-Swallow and Haksar, 2019). In the case of personal data, the person whose information details have been recorded, is referred to as the data subject. A data collector collects and stores data. In doing so, the data collector incurs costs. On the demand side, the data processor uses data and aggregates and analyzes them. In practice, the data collector and data processor could be the same organization.

2.3.2 The data economy

More and more economic activities take place within the data value chain. Those activities and the connected supply chains are thereby becoming more important parts of the overall economy. To monitor the impact of the data economy, the European Commission uses the following definition[2]

The Data Economy measures the overall impacts of the Data Market on the economy as a whole. It involves the generation, collection, storage, processing, distribution, analysis elaboration, delivery, and exploitation of data enabled by digital technologies. The Data Economy also includes the direct, indirect, and induced effects of the Data Market on the economy

Using this definition, the size of the data economy in 2019 was estimated to be 2.6% of GDP for the EU (325 billion euro, excluding UK). Moreover, the data economy is expanding rapidly. In a conservative scenario the data economy is forecasted to grow to 430 billion euro in 2025 (3.3% GDP), while in the most aggressive outlook its size is forecasted to become 827 billion euro by 2025 (5.9% of GDP).[3] In a recent complementary effort to define the size of the digital economy, the OECD stresses that there 'remains some subjectivity or "fuzziness"' in turning definitions into numbers (OECD, 2020). Thus, the absolute numbers of these estimates depend on how the definition is translated in practice and are therefore somewhat arbitrary.

[2] See for instance the European Commission communication on "Building a European data economy" (link)
[3] "The European data market monitoring tool", 2020, Directorate-General for Communications Networks, Content and Technology (link)

3 Market failures in the data economy

3.1 Externalities

Externalities in the data economy occur when data transactions impact third parties. MacCarthy (2010) introduced the concept of privacy externalities: people disclosing information might reveal information about others. This information can be revealed directly, as is the case with pictures or conversations concerning multiple people, or inferred by leveraging data analytics. Some striking examples include the disclosure of sexual orientation[4] and ethnicity[5]. The Cambridge Analytica scandal serves as the most well-known example of how social media profiles can be used to infer (political) preferences and can be exploited to influence actions (such as voting behavior and ultimately elections). In the case of Cambridge Analytica the extent of the effect of social media profiling is still a subject of scientific debate.

Economists have recently begun to formalize the concept of privacy externalities. Choi et al. (2019) developed a theoretical privacy model in which firms collect data that require consumers' consent. Privacy externalities cause excessive data collection in a monopoly setting. They conclude that 'the current main privacy regulatory framework of the informed consent model may be ineffective to address the privacy concerns associated with the data broker industry' (p. 122).

Acemoglu et al. (forthcoming) study the impact of privacy externalities on the "price" of data. Consumers typically do not receive money for transferring their data, but are paid in the form of getting access to a service, such as a search engine or a social network. The "price" of data is the utility value level of those services. In their model, the price of data is suppressed due to externalities and consumers choose to share an inefficiently high amount of data. To understand how this works, consider two persons who value privacy differently and who are potential customers of a digital platform. The first person is willing to transfer data to the platform because the utility he receives is higher than the cost incurred due to his reduced privacy. The second person on the other hand values her privacy more than the use of the platform: she thus refrains from using the platform. Things change however when the data of the first person reveals information about the second person, i.e. when privacy externalities are introduced. If the privacy of the second person is significantly impacted by the choice of the first person to join the platform, there is little reason for her to abstain from using the platform. She will thus join the platform but the value of her own data is reduced since some information was already captured. Consequently, the outcome of this game features excessive data sharing, excessively low prices for data and a reduced consumer welfare level.

Privacy externalities can lead to social welfare decreasing activities (Bergemann and Bonatti, 2019; Bergemann et al., 2020). For example, privacy externalities facilitate third degree price discrimination[6]. That is, the externality can be leveraged by segmenting users into groups and discriminate these groups in terms of price or product offerings. Privacy externalities make it worthwhile for firms to collect more data from an individual as this data can be extrapolated to a larger customer base. There is a risk that companies then incentivize users

[4] See e.g. Jernigan, C., & Mistree, B. F., 2009, Gaydar: Facebook friendships expose sexual orientation. *First Monday*, 14(10).
[5] See e.g. Annalee Newitz, 2016, Facebook's Ad Platform Now Guesses at Your Race Based on Your Behavior, *Ars technical* (link)
[6] Whether third degree price discrimination leads to an increase or decrease in social welfare depends on the shape of the demand function. For linear demand functions, it is welfare decreasing.

to spend ever more time on platforms beyond the benefit of the individual user. Bergemann and Bonatti provide an intuitive explanation for the importance of externalities in the data economy by considering the counterfactual: what would happen when data from individuals would not contain any information about others? They postulate that it would become *less* interesting for firms to gather data and secondly that individuals would be *less* willing to give up their data as their bargaining power increases.

A second type of externality occurs when data in a transaction is available to third parties. E.g. via APIs or public databases. In that case, other parties are able to use the data to their own benefit. Take Twitter data for example. It is relatively straightforward to scrape tweets and this information can then be leveraged to improve customer sentiment tooling. We stress that this positive externality, which enables others to reap the benefits of the nonrivalry of data, only exists when data is made non-excludable.

Jones and Tonetti (2020) built a theoretical framework to understand how property rights impact the potential positive externalities of data. First, to include privacy considerations, they assume that consumers incur a utility cost when firms own their data. Second, their model shows that social gains materialize when data is used by different firms simultaneously. In other words, when data is *not* shared between firms, there is an economic inefficiency because the nonrivalry of data is effectively ignored. Shifting property rights from firms to consumers leads to an outcome which is closer to the social optimum as consumers balance their privacy needs with economic gains. In their paper, Jones and Tonetti explicitly depart from the Coase theorem. According to this theorem the initial allocation of property rights is irrelevant as long as transaction costs are zero and property rights are clearly defined. With these conditions in place property rights redistribute to the party who values them the most. For data, however, the prerequisites are not upheld and the initial allocation of property or usage rights matters. First, as Jones and Tonetti lay out, when property rights initially sit with consumers, it is unrealistic to expect that they will exclusively trade their data with one firm. In fact because of the nonrivalry of data the opposite seems to be the case: consumers trade their data with different firms simultaneously. Second, we note that due to privacy externalities and the fact that transactional data often involves multiple parties defining initial data property rights is challenging.

3.2 Public goods

Due to the nonrivalry of data, it can either be classified as a public good or a club good depending on its excludability. Non-exclusive, nonrival goods are public goods. Provision of public goods is not a given. Often the free-rider-problem occurs: people or organizations benefit from using the good while underpaying for the maintenance/ creation of the good. Examples of public goods include clean air, knowledge and street lighting.

Data as a public good already exists. Prime examples of institutes setup to organize data as a public good are the national statistics offices around the globe (Carrière-Swallow and Haksar, 2019). Similarly, researchers have put together databases that strengthen various scientific activities, such as the Protein Databank and the Human Genome Project (Hill et al., 2020). Biobanks in which biological samples are stored facilitate medical research and help detect patterns in populations, e.g. during pandemics.

The role of adequate public good provision in the data economy is not limited to data itself, but extends to the concept of privacy. Fairfield and Engel (2015) argue that privacy should be seen as a public good due to the negative externalities that we discussed earlier. An adequate protection of privacy requires collective action in a similar way that public goods do. In their view, privacy regulations should therefore shift from being solely focused on the individual towards empowerment of groups. They discuss in depth how insights from behavioral economics can be leveraged to overcome current ineffective policies. They conclude that:

'Tools should not be centered on individual rights of review and deletion, which have proven largely ineffective. Rather, tools should focus on group communication, sanction, and fostering a sense of repeat play and community. Even the way that we speak about the nature of the problem can have an impact on whether people cooperate to produce the public good of privacy.' (p. 457)

3.3 Market power

The data economy is characterized by large tech firms that dominate the market for consumer data. In Europe and the United States, Google is by far the largest player when it comes to search and web browsing, while Facebook is leading in social media applications. Kirpalani and Philippon (2020) analyze theoretically how data sharing by consumers impact the market power of two-sided platforms (such as Amazon or in some instances Google). In their model merchants and buyers (i.e. consumers) interact via a digital platform. They conclude that from a social planner standpoint consumers share too much data with platforms as it leads to less competition for two reasons. First, because of their access to consumer data and therefore their understanding of consumer preferences, platforms are in a good position to become sellers themselves. Secondly, the outside option for merchants, i.e. to sell their products outside the platform, becomes less appealing as the platform helps in matching supply and demand. Both effects lead to an increase in bargaining power of the platforms relative to the merchants and ultimately to lower consumer welfare.

The Furman report (2019) sums up several reasons that explain why concentration in the data economy is likely to occur. First, digital applications display strong economies of scale and scope. Initial investments are high, but marginal costs are close to zero. Once a successful digital product has been built, it can quickly be rolled out globally. Sharing consumer data internally allows a firm to expand its scope relatively easily into adjacent markets. Second, data acts as a barrier to market entry. Incumbents have a competitive advantage as they can leverage incoming and historic data to improve product offerings. With these improvements they are then likely to attract new customers and thereby create a positive feedback loop. Third, digital platforms often feature network effects. A network effect occurs when the use of a platform becomes more valuable when the number of users increases. Fourth, there may be significant switching costs involved that prevent users from leaving incumbent firms. E.g. when reputation data is tied to the use of a particular platform. Lastly, it may be difficult for new entrants to obtain investments necessary to build intangible capital.

Crémer et al. (2019) analyze in depth how market power in the data economy impacts competition policy. One of their main conclusions is that access to data needs to be taken into account in antitrust assessments. Furthermore given the natural tendency towards concentration for digital products and services, assuring that competition *for* the market can take place becomes key.

3.4 Information asymmetry

Information asymmetry refers to situations where one party in an economic transaction has more information than the other party. One party could be better informed about an existing situation, such as the quality or coverage of existing data. Economists refer to this type as ex ante asymmetric information or adverse selection. Another form of information asymmetry is where one party is better informed about the behavior of the other party. In the context of the data economy this could refer to uncertainty about the type of data analysis that is conducted or the effort the other party exerts in data protection. This form of asymmetry is called ex post information asymmetry or moral hazard.

The data economy witnesses both increasing and decreasing levels of information asymmetries depending on the context. To see how the use of data can lead to reduced asymmetries, consider the above example of moral hazard in insurance. The risk of moral hazard reduces when insurers are able to monitor behavior and adjust fees accordingly. Such schemes are actively being developed by insurers worldwide. Car insurers for example offer products in which driving style is monitored and taken into account by adjusting fees, so called "pay-how-you-drive auto insurance" or PHYD. Reimers and Shiller (2019) analyze PHYD in the US and they find a meaningful negative effect on fatal car accidents. Combining data about loans from a variety of financial institutions by credit agencies is another example in which data is used to decrease the information asymmetry between consumers and firms. This reduction improves the functioning of credit markets (Carrière-Swallow and Haksar, 2019).

At the same time, the information gap between consumers and firms that use their data is often widening. Technology firms are more and more capable to accurately predict consumer behavior by rigorously analyzing large, combined, data sets. Amazon for example recently patented a business method through which it is able to ship products *before* customers ordered[7]. Thus in some instances tech firms know more about their customers than customers know about themselves (Zuboff, 2015). A future development could be that insurance firms know more about their clients' health via access to DNA data or smart health devices.

Efforts to overcome the information asymmetry have relied on the *informed consent* model. In this approach, data users (e.g. firms) inform data providers (e.g. consumers) how their data will be used before the two parties agree to transfer data. The leading ideas behind this model are that 1) privacy is an individual matter, and 2) consumers are able to make rational decisions about data transactions once they are informed by the data processor on how their data is used. Unclear is the economic rationale of the informed consent model; firms may lie or misrepresent what they will do with the data, which raises doubts about the credibility of their claims. In practice, this model turns out to be problematic (MacCarthy, 2010; Zuboff, 2015; Wachter and Mittelstadt, 2019). The sheer number and length of privacy agreements leads to information overload. Privacy agreements are therefore hardly read (Obar and Oeldorf-Hirsch, 2020)[8]. It is likely also naïve to expect that the complexities of data processing and analysis can be fully understood by consumers in the first place.

3.5 Behavioral biases

Throughout the data value chain behavioral biases appear that hinder fair data transactions. Acquisti et al. (2015) therefore argue that 'To be effective, privacy policy should protect real people—who are naïve, uncertain, and vulnerable— and should be sufficiently flexible to evolve with the emerging unpredictable complexities of the information age.' (p. 514) Their view is based on an extensive review of behavioral studies that indicate people's privacy preferences are often 1) uncertain, 2) context dependent and 3) malleable. Consumers' uncertainty about privacy preferences implies that consumers are not always able to apply a rational cost-benefit analysis when their privacy is at stake. E.g. present-bias creeps in that causes consumers to overvalue short-term gratification over the long-term consequences of privacy loss. How much data consumers are willing to give up depends on the context. E.g., people are more likely to share data when they see that their friends or colleagues are doing the same. Lastly, the presence of behavioral biases also means that organizations can manipulate consumer behavior to their own benefit. Websites already leverage biases to persuade consumers to hand over data (Smith et al., 2013) or to buy products[9]. Booking websites and online retailers are for example well known to suggest scarcity to increase sales.

[7] See e.g. this link
[8] In 2008 researchers from Carnegie-Mellon showed that reading privacy-agreements fully would lead to 781 billion dollar in costs for the US economy: McDonald, A.M. and L.F. Cranor, 2008, The cost of reading privacy policies, *Isjlp*, vol. 4, pag. 543.
[9] See this blogpost for examples.

4 Policy pathways for the data economy

In our Policy Brief we describe three pathways for policy makers to strengthen the data economy. The first pathway explores ways to improve policies that focus on individual decision making. Relying on the *informed consent* model is reasonable when externalities and the difference in bargaining power are small. However, when externalities or the difference in bargaining power grow larger it becomes worthwhile to shift focus from individual decision making to collective action. Lastly, imposing and enforcing rules by governments or agencies constitutes a third pathway that is especially relevant when externalities are very large or the distribution of bargaining power is out of balance. In this chapter, we provide more details about the policy options within these pathways and, where possible, make a connection with the economics literature. Table 1 summarizes different policy options as discussed in this document. We also review some of the recently proposed EU legislation for the data economy in Box 1.

4.1 Improving individual decision making for data sharing

In our analysis so far, we have come across several reasons for why the current policy focus of informed consent is at best incomplete. In this section we ignore some of the underlying difficulties on purpose: it is assumed externalities are relatively small and bargaining power between data subject and data holder are in balance. These conditions make it more likely that informed consent is a viable policy option. Under these circumstances policies that primarily aim to respect individual privacy preferences as much as possible seem justified. The main hurdles to overcome are then related to ensuring that 1) information provided upfront is easily accessible and understandable, and 2) that the choices at hand are meaningful and relevant in the specific use case.

4.1.1 Improving information provision

Currently, most data transactions under the informed consent model take place by asking consumers to sign a privacy agreement. The idea is that these agreements serve as the necessary notice that informs data subjects about the use of their data. The information transfer via these agreements is flawed for three reasons. First, privacy policies are lengthy and difficult to understand. Second, the sheer number of interactions that require personal data typically leads to a bombardment of privacy agreements. Third, combining data makes it difficult to grasp how revealing an isolated piece of data can be (Susser, 2019).

From these flaws, several improvements for information provision under the informed consent model follow. Encouraging the use of privacy certificates could build trust among consumers and bypass the need to dive into privacy agreement details (Bijlsma et al., 2014). National privacy authorities or certified private parties are both potentially suitable parties for providing these type of certificates. Standardization and a way to accept multiple privacy agreements at once could prevent an overload of individual privacy agreements. Such standardization could for example materialize via browser controls. Part of this standardization could entail a condition on not recombining data with other sources.

Table 1: overview of several policy options for the data economy

	Policy instrument	Policy legitimation	Risks and challenges
Strengthening individual decision making	**Privacy certificates**	Reducing information asymmetries	• Potentially ineffective without meaningful alternatives for consumers
	Transparent and standard privacy agreements	Lowering transaction costs and reducing information asymmetries	• Potentially ineffective without meaningful alternatives for consumers
	Stronger collaboration (incl. data sharing) between relevant regulatory authorities	Improving effectiveness of regulatory bodies given that market power, privacy and innovation are increasingly tied in data economy	• Shift of tasks/ power from legislative power to regulatory authorities • Privacy risks when data get shared • Need for supranational collaboration
	Facilitating data sharing coalitions	Lowering coordination failures	Benefits likely to be unevenly distributed amongst participants (e.g. firms with strong data analysis skills gain more)
Collective action	**Collective data agreements**	Internalizing external effects and balancing negotiation power	• Individual rights are jeopardized • Potentially difficult to form collectives
	Creating incentives for data sharing	Realizing positive external effects	Reduced incentive to collect data
Direct government intervention	**Decentralized data storage**	• Increasing negotiation power consumers • Improving security of personal data	• Costs likely to shift to consumers • Network effect of applications might remain and reduce impact
	Restrictions for collecting specific data types	• Reducing negative externalities	• Level of innovation might decrease
	Creating public databases	Enabling positive externalities	Lower incentive for individual organizations to collect data/ risk of free riding
	Mandatory data unbundling	• Increasing negotiation power consumers • Decreasing negative externalities	• Uncertain impact on business models • Potentially reduced levels of innovation • Potentially more barriers for access to digital products and services
	Requirements to offer alternative products or services (e.g. via do-not-track registers)	• Increasing negotiation power consumers • Decreasing negative externalities	• Potentially reduced levels of innovation • Potentially more barriers for access to digital products and services

4.1.2 Creating meaningful choices

A major critique on the *informed consent* model concerns the absence of a meaningful choice. Often accepting privacy agreements are a prerequisite for accessing websites or getting full functionality. In some cases, the limited choice is connected to the underlying business model. For instance, targeted advertising is an important way to generate income. Shifting towards subscription based business models might remove the need for companies to collect large amounts of data and provide alternatives for "free" products. Governments may in some cases decide to force companies to do this. By ensuring ample competition, e.g. via antitrust regulation that remains relevant in the digital era (Crémer et al., 2019), governments may also create market conditions that foster entry of companies that offer meaningful alternatives.

4.1.3 Strengthening regulatory oversight

Stronger coordination between regulatory bodies is needed to ensure effective oversight in the data economy. Data start to become key in the work of multiple supervising bodies that were setup to safeguard the common good in economies. Privacy, consumer protection and market competition increasingly depend on how data are used and shared in the data value chain. The scope of regulators therefore starts to overlap[10]. To protect the consumer effectively in the data economy, therefore requires stronger coordination and cooperation between regulatory authorities, also across borders. This could, e.g., include the sharing of relevant data.

4.2 Creating collective action in the data economy

In this section we focus on instruments for collective action. First, the concept of collective data agreements is put forward as a way to internalize externalities and bundle bargaining power. We look into similar concepts derived from common pool resources to understand potential pitfalls. Second, economic incentives for data sharing are discussed as well as different ways to organize data sharing.

4.2.1 Collective data agreements

The Policy Brief recommends policymakers to explore the policy option of allowing people and organizations to collectively bargain about the use of personal data and settle on collective data agreements (cdas). This policy option allows people and organizations to voluntarily move away from an individual "informed consent" data management system and find a mutually advantageous solution. A more radical policy option is to legally require firms in specific situations to enter negotiations with people about the use of their data.

Though cdas have not, to the best of our knowledge, been used in the real world so far, the idea of a more collective approach to data contracts has been suggested by several researchers and data experts. Moreover, cdas have certain aspects that are studied in the economics literature. This section discusses other collective bargaining solutions and gives an overview of the earlier literature on collective data solutions.

[10] See for instance this report by Aline Blankertz that points out how privacy and competition are related

Notes on EU regulation (1/2)

In the second half of 2020, the European Commission put forward three legislative proposals to promote the common good in the data economy: the Data Governance Act, the Digital Markets Act and the Digital Services Act. In this box we discuss some of the proposed policies and how they relate to our previous discussion on market failures in the data economy.

The **Data Governance Act** (DGA) is the first out of multiple instruments announced in the 2020 European strategy for data. The Act is set to encourage the creation of infrastructure for data sharing and to create EU-wide data spaces in strategic sectors such as healthcare, energy and mobility. The measures aim to promote the positive externalities that arise when data becomes more accessible. The DGA leans on four pillars. Firstly, the DGA sets conditions for reuse of public sector data that is subject to existing protection (such as commercial confidentiality, intellectual property or data protection). Secondly, the Act establishes a framework for new data intermediaries – entities that provide various types of intermediation services. These entities could play a role in making collective data agreements a possibility as it foresees the creation of 'data cooperatives' that 'strengthen the position of individuals in making informed choices before consenting to data use'. Thirdly, the Act introduces the concept of data altruism meaning that individuals are encouraged to voluntarily donate personal data to serve the common good. Fourthly, the Act establishes a European Data Innovation Board, a formal expert group which will coordinate national practices and policies on data sharing and will oversee data intermediaries.

The main goals of the **Digital Services Act** (DSA) are to 1) safeguard consumers and their fundamental rights, 2) create clarity on what is expected of online intermediaries in terms of accountability and transparency, 3) stimulate innovation and ensure competitiveness within the EU. The rules are aimed at online intermediaries, which e.g. include hosting and cloud services, and online platforms (such as online marketplaces and app stores). The proposal explicitly mentions that measures are asymmetric. Special, stricter, rules apply to the larger online platforms as the potential societal harm is assumed to increase with platform size.

Some rules in the proposal reinforce the informed consent model. E.g., the DSA includes measures to improve the transparency around content moderation and requests larger platforms to provide meaningful information about targeted advertisements. Other parts of the proposal are supposed to reduce the burden on small firms to ensure compliancy, e.g. by creating standards and guidelines. Large platforms are obliged to grant data access to researchers who aim to study systemic risks associated with the use of platforms (such as virality of messages).

The proposed regulation within the Digital Services Act reflects some of the market failures in the data economy. First, as the measures are geared towards the largest platforms, the proposal potentially results into a more balanced distribution of market power. Second, the measures address information asymmetries between consumers and firms. Third, by having access to data independent researchers might better understand the risk of privacy externalities within different platforms.

Notes on EU regulation (2/2)

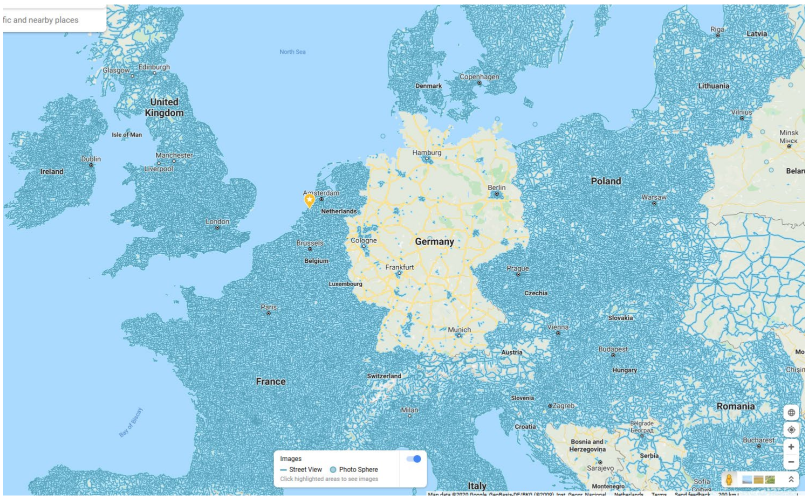
The **Digital Markets Act** (DMA) enables the European Commission to impose behavioral restrictions and obligations on specific large online platforms – so called gatekeepers. A firm can be qualified by the Commission as a gatekeeper when the firm has a strong and durable economic position, is active in multiple EU countries and links a large user base to a large number of businesses via an online platform. Examples of rules for gatekeepers are to provide access to data of business users that those businesses generate on the gatekeeper's platform and not to engage in self-preferencing by treating its own products more favorably in ranking than similar products offered by third parties on the platform. Another interesting requirement is that gatekeepers may not combine personal data that originates from the platform with data from other services that the gatekeeper offers or personal data purchased from other firms. This requirement can be interpreted as 'ringfencing'. The DMA can be seen as a digital complement to the existing general EU competition law, as it enables the Commission to ex ante regulate platforms – before the platform even abuses its dominant position. A possible rationale for the DMA is that it allows the Commission to intervene quicker than via the long trajectory of an abuse of dominance case.

Similar to the DSA, the DMA seems to have the potential of addressing problems of market power and unbalanced bargaining power, because it applies to large platforms and is designed to increase the position of smaller firms that are dependent on those platforms. The ringfencing requirement prohibits the combination of personal data and therefore limits the incentives of platforms to collect data. This may lower the risk of negative external effects. However, the DMA provides gatekeepers with an option to circumvent the ringfencing requirement when they have an informed consent from the end user.

Examples of collective bargaining agreements

Perhaps a precursor to a cda is the reception of Google Street View in Germany. Street View offers users to see and navigate panoramic photos at street level. It is a complementary service to Google's map service Maps. In much of Europe, Google has managed to photograph the continents' streets, see Figure 2. Although individuals have, under GDPR, the right to request Google to blur photos of their property, most people apparently do not use this option. An exception is Germany, where large parts of the country remain uncovered (see also Zuboff (2019)). In Germany, many citizens requested Google to blur their house or other property ('verpixeln'), which increased the costs for Google to maintain a Street View service there. After a public debate and a denunciation of Street View by the vice-chancellor in 2010, Google decided to largely suspend the service in Germany.

Figure 2 Partial coverage of Google Street View in European countries (screenshot taken in 2020)



The Street View case holds several lessons for collective data agreements. First, the case illustrates that data processing firms, such as Google, have much bargaining power over individuals. Secondly, individual consent does not change much to this imbalance, because most inhabitants do not care enough or are unaware of negative consequences. In the context of Street View, possible negative consequences are loss of privacy or a higher burglary risk. Thirdly, the case illustrates that collectively, people can force data processing firms to choose another course of action. Finally, both in Germany and in Street View-friendly countries, there seems to be an absence of collective decision-making. This makes it unclear whether the German outcome really is preferable to most Germans. Many people may likewise benefit from the availability of Street View and the public uproar against Google prevents them from enjoying these benefits. This suggests that a structured collective decision mechanism increases the odds that the bargaining outcome strikes the right balance.

Outside the realm of data and digital markets, there are several other areas where people have chosen to collectively organize to solve a shared problem. In labor markets, workers organize themselves in labor unions. The traditional role of labor unions is to enable workers to collectively bargain with firms over work conditions and payments. The firm-level and aggregate effects of labor unions are central topics in labor economics and the literature does not yet seem to converge to a consensus view. Notable contributions are Blanchard and Summers (1986), who develop a model that links the relatively high levels of unemployment in Europe to the labor union penetration, and Card (1996) who empirically establishes a positive wage effect of unions for low-skilled workers. More recently, studies by Benmelech et al. (2020) and Azar et al. (2019) document a positive correlation between wages and employer concentration (measured with the Herfindahl-Hirschman index, which could indicate bargaining power on the side of firms).

Another example of collective solutions are homeowner associations (hoas). These associations hire privately managed property companies to solve the problem of collective action in privately owned neighborhoods. Drawing on experience from Taiwan, Chen and Webster (2005), argue that hoas are a more effective solution than direct government involvement or voluntary individual action. Chen and Webster note that hoas still suffer from problems such as information asymmetry or rent-seeking. To foster hoas, governments need to provide appropriate enabling legislation. In the Netherlands, apartment owners are legally obliged to form a hoa, with the objective of maintaining common property, such as the roof, the building’s façade or the common staircase. These legal requirements help individual apartment owners to solve the problem of the provision of a public good.

The literature on common pool governance also proves relevant for cdas. Because of nonrivalry, data seem almost the opposite of a common pool. Privacy however, can be seen as a 'commons', because when someone decides to share information with a firm, this transaction tends to leak data about others who are in the same social circle. The common characteristic of data markets and common pools is that individual choices have external effects. In other words: if I decide to fish more, there will be less fish left for you and when I decide to share data, there will be less privacy for you.

In settings with external effects, economists typically offer two solutions: defining property rights in the spirit of Coase (1960) or 'Pigovian' corrective taxation (Pigou, 1920). Those approaches are not realistic in the data economy, for reasons explained above. In her pathbreaking work on the management of common pool resources, Ostrom (1990) points to a third way. Ostrom (1990) documents that in real world commons, ranging from communal meadows, forests, to irrigation reservoirs and ditches, existing communities often manage to develop institutions that foster cooperation. Ostrom (1990) concludes that the solutions people find differ across settings and seem to be tailored to the specific environment. In other words: there is no one-size-fits-all mechanism that works in every situation. She identifies, however, eight distinct common properties of successful institutions. These common properties, or so-called 'Design Principles Illustrated by Long-Enduring Common Pool Resource Institutions' are (the following list draws on Coyle et al., 2020):

1) There are clear boundaries and rules about who is entitled to what.
2) Monitoring actions is feasible.
3) There are mechanisms for resolving conflicts.
4) Individual responsibilities and benefits broadly balance.
5) Users themselves are responsible for monitoring and enforcement.
6) Sanctions for abuse are possible and graduated, getting progressively tougher.
7) Decisions are legitimated by the participation of users.
8) Decisions are also legitimated by government recognition.

Let's look at Ostrom's principles through the lens of a collective data agreement. The first principle suggests that in a cda, the agreement should define and clarify how data will be shared, stored, accessed and monetized throughout the entire data value chain. The second principle implies that the cda should be transparent and auditable. The third principle can be supported by legislation that appoints a regulator or authority that can act as an institution for conflict resolution. To ascertain the fourth principle, of a balanced outcome, it is vital that all relevant stakeholders are part of the cda and that the ones who represent the data providers have sufficient bargaining power. The fifth principle seems harder to fulfill in the data economy. A possible solution is that the data providers (the consumers) delegate monitoring and enforcement to a specialist, perhaps an authority. The cda can, to be consistent with the sixth principle, also describe the sanctions for abuse, for instance by specifying a fine or another consequence. The penultimate principle is difficult to impose top-down, but can develop as experience with cdas accumulates. As users and firms see the advantages of a cda they are more likely to participate. Finally, the eighth principle may need some enabling legislation. In particular, there may be a need for a legal exemption from GDPR for parties who voluntarily choose to participate in a cda.

The risk of free-riding is of critical concern in reaching stable collective bargaining outcomes. In international environmental cooperation, for instance, countries value a clean environment, but prefer other countries to bear the costs. Barrett (2016) shows in a simple game-theoretic model how institutions such as treaties help countries achieve better outcomes. Cooperation may also be better sustained when agents interact more frequently, such that, formally, the game changes from a "one-shot" game to a "repeated game" and agents can coordinate on a more efficient outcome. These insights might also help in reaching successful collective agreements in the data economy.

Literature on collective data agreements

Individual consent, one of the pillars of current privacy regulation, has drawbacks when externalities may arise or when individuals are behaviorally biased. These notions are not new and have been addressed by several authors. In an early contribution, Cassell and Young (2002) discuss the shortcomings of the individual consent approach in the context of health care research and argue that informed consent should not be prioritized above other public interests. The informed consent model is thoroughly criticized by MacCarthy (2010), who seems to be the first to argue that negative privacy externalities arise when individuals share personal data. Acquisti et al. (2015) provide an overview of limited rational behavior of consumers in the context of privacy decisions. Their reading of the evidence is that privacy preferences and behaviors are malleable, because they are largely context-dependent. As a result, firms can change privacy behavior by manipulating the context. Wachter and Mittelstadt (2019) criticize GDPR on the grounds of what they see as an inability of the regulation to protect individuals against big data inferences that could damage one's privacy or reputation. Their proposed solution is a 'right to reasonable inferences'.

Several authors have argued, for a variety of reasons, to complement, or even replace, individual consent with a more collective consent model. Similar to Wachter and Mittelstadt, Mantelero (2016) argues that a shift is warranted from individual to collective data protection rights, motivated by the advent of big data analytics. Bygrave and Schartum (2009) introduce the notion of 'collective consent', which means 'consent exercised on behalf of a group of data subjects, but without these persons individually approving each specific exercise of the decisional competence.' This notion is a close analog to our concept of a cda. An advantage of collective consent[11] over individual consent is, according to Bygrave and Schartum (2009), that it enables data subjects to improve their bargaining power over data controllers and that it reduces the transaction costs that are associated with time-consuming processes of obtaining consent on an individual basis. In a blog post, Jeni Tennison[12] perceives individual informed consent as 'fundamentally broken', because of limited rationality and external effects. Akin to Bygrave and Schartum's concept of collective consent, Tennison coins the term 'community consent', which is, ultimately, an obligation to data processors to 'provide evidence that their use of data is acceptable to those affected by it, and to important subgroups that may be differentially affected'.

4.2.2 Incentives for data sharing

Economic rationale for data sharing

In the Policy Brief, one of the policy directions deals with creating more incentives for data sharing. Another policy option considers the creation of public databases. Organizations would then be obliged to provide data for the common good[13]. Widening the access to data allows others to make use of the same data. The rationale for public investments in data sharing is to enable a positive externality. In the words of the OECD: 'data sharing may benefit others more than it benefits the data creator and controller, who cannot privatize these benefits and as a result may not sufficiently invest in data sharing or may even refrain completely' (OECD 2015).

Hill et al. (2020) use the example of economics research to demonstrate that more effort towards building a common data infrastructure would be worthwhile. They argue that the lack of high quality public datasets in economics is very similar to a classic public good provision problem. Researchers are incentivized to leverage a dataset primarily for their own research because by doing so they safeguard their own publications, instead of

[11] See also this blogpost by Anouk Ruhaak for a discussion on collective consent
[12] See this link.
[13] In a recent external study published by the Dutch ministry of Economic Affairs several other schemes for mandatory data sharing in the technology sector are considered (link).

generalizing a dataset in such a way that it optimizes social benefit[14]. As a consequence, researchers spend many hours redoing the work of others.

As Hill et al. (2020) point out, other scientific disciplines have used a variety of methods to overcome this public good provision problem. In observational astronomy, central organizations are responsible for setting up the infrastructure for data collection via telescopes. Researchers are then granted access to these telescopes on the condition that any data collected will become publicly available afterwards. Structural biologists are required to deposit data on new biological structures into a public database prior to publication of scientific articles. The Human Genome Project is a third example where scientists collaborate to maximize the benefit from data sharing. Centralized institutions can play an important role in creating both the right governance for data and assure data quality.

The social benefit of sharing data varies on a case-by-case basis. Studies on other centralized forms of information sharing suggest that returns can be significant. Furman and Stern (2011) econometrically analyzed the impact of a biological research center that certifies and disseminates knowledge. Using a difference-in-differences method, they find that knowledge accumulation increases due to the efforts of this center. In 2008 the United States Geological Survey made its Landsat data publicly available. Several studies indicate that this has led to new applications and scientific research (Loomis et al., 2015; Zhu et al., 2019; or Nagaraj, 2021). It has for example been used to quantify deforestation and finding water bodies. Societal benefits (in the order of billion dollars) far exceed the initial investments. In 2013, McKinsey estimated that open data yields a 3 trillion dollar economic opportunity globally.[15]

Different ways to organize data sharing

There are different ways to create a sharing data infrastructure. Researchers from New York University have developed a typology to clarify the different institutional arrangements and operational dynamics that enable data sharing.[16] Their data collaborative matrix uses two variables, engagement and accessibility, to define six types of data collaboration, see Table 2.

Table 2: Typology of data collaboratives

	Open access	Restricted access
Independent use	Public interfaces	Trusted intermediary
Cooperative use	Data pooling	Research and analysis partnership
Directed use	Prizes & challenges	Intelligence generation

Source: www.datacollaboratives.org

Public interfaces arise when individual organizations provide open access to their data, for instance by using Application Programming Interfaces (APIs). Firms can also collectively decide to give access to unified data and thereby create data pools. In the case of prizes and challenges access to data is made available with a specific goal in mind such as predicting heart diseases or improving self-driving car algorithms.

Organizations can also decide to provide restricted access to their data. Trusted intermediaries allow firms to collaborate with data users while upholding strict access control. In some cases analyses following the restricted access are made public. Firms can also partner together with researchers to create new knowledge. A good example is the Dutch Consumer Price Index for which supermarket chains provide data to Statistics

[14] There are some exceptions to this rule. Examples include the Penn World Table (link) and the Maddison project (link).
[15] McKinsey, 2013, Open data: Unlocking innovation and performance with liquid information (link)
[16] GovLab, 2019, Leveraging private data for public good (link)

Netherlands. This allows for a better measurement of the level of inflation. Lastly, firms can choose to share insights generated from their data (intelligence generation).

The above typology illustrates that there is no one-size-fits-all model for data sharing. It requires sector specific analysis to understand which type of data collaborative offers the highest social returns. Knowledge on the different types of data sharing is rapidly increasing. The Open Data Initiative has for example put considerable research efforts into public interfaces (open standards and APIs) and data pools/trusted intermediaries.[17]

Moreover, a careful governance design is needed to make any type of data sharing effective. Koutrompis, Leiponen and Thomas (2020) discuss in more detail which factors are critical for successful data marketplaces. They stress that for a successful marketplace to emerge rigorous provenance is key. Metadata on the origin, content and collection methodology as well as clear user rights are therefore important. Provenance can become particularly problematic for large multilateral data platforms. Here, the concept of a data trust might be useful in which a separate legal structure is tasked with providing independent stewardship of data.[18]

Concrete opportunities to create value by data sharing

The aforementioned McKinsey report lists opportunities for value creation by data sharing across sectors. In Figure 3, we summarize the key potential applications per sector as listed in the report[19]. To get a feeling for some of the opportunities we discuss two applications in more detail.

Figure 3 Key potential applications for data sharing by sector

Education	Transportation	Consumer products	Electricity	Oil and gas	Health care
• Improved instruction • Matching students to programs • Matching students to employment • Transparent education financing • Efficient system administration	• Improved infrastructure planning and management • Optimized fleet investment and management • Better-informed customer decision making	• Improved product design and manufacturing • Efficient store operations • More targeted marketing and sales • Better-informed consumption • Improved post-sales interaction	• Optimized generation investment • Efficient generation operations • Optimized investment in transmission and distribution • Efficient transmission and distribution operations • Optimized retail and consumption	• Optimized upstream investment • Efficient upstream operations • Optimized midstream and downstream investment • Efficient midstream and downstream operations • Better-informed consumption	• Right living • Right care • Right provider • Right value • Right innovation

Source: McKinsey (2013)

In the transportation sector, data on passenger flow and travel times can help build better infrastructure and improve operations (see for instance O'brien et al. (2014) for an example based on bike sharing) . Relevant data sits with governmental bodies (e.g. demographic data), companies (e.g. data from public transport operators on vehicle occupation or real-time vehicle location), consumers (e.g. data on travel times, preferred transportation modes) and third-party-providers (e.g. data on specific events). An example of leveraging data sharing for mobility solutions is oneTRANSPORT in the United Kingdom.[20] This data marketplace has allowed for managing traffic flow during major events and providing information to travelers entering city centers.

[17] See this link for their work on open standards and APIs. The ODI has also written a lot about data trusts. A legal structure to organize data pools or trusted intermediaries (link).
[18] See the discussion on data trusts from the Open Data Institute (link)
[19] We have only included those sectors for which the impact has been quantified. The McKinsey report also gives examples for data sharing in the domain of consumer finance.
[20] See their website (link) and the case study discussion by the Royal Academy of Engineering (link)

Data sharing in education can spur innovative personalized learning methods. Acemoglu and Restrepo (2020) have argued that leveraging artificial intelligence (AI) effectively to create adaptive education can increase productivity within classrooms: 'AI software can be designed to collect and process in real-time data about the specific reactions, difficulties and successes students have in different subject areas, especially when taught in different styles, and then make recommendations for improved individualized teaching.' Building such personalized learning solutions would require data from students, schools and testing agencies.

4.3 Restricting parts of the data value chain

So far we have focused on policy options that either facilitate individual decision making or options that instigate collective action. In this final section, we will instead briefly describe three other policy options for the data economy where the government intervenes directly by creating restrictions in parts of the value chain. There are potentially many different ways for governments to intervene directly. Here, our aim is not be complete, but rather to provide an outlook that could inspire further research or pilots.

4.3.1 Collection of data: encouraging decentralization of data storage

In a 2018 blogpost Tim Berners-Lee, the inventor of the world wide web, wrote that 'the web has evolved into an engine of inequity and division; swayed by powerful forces who use it for their own agendas.'[21] To rebalance powers, he is developing a platform (solid) to give 'every one of us complete control over data, personal or not'. The main idea behind this decentralized platform is to provide consumers with a personal data space that contains all their private and public data. Organizations can ask for permission to access the data in order to run applications (Mansour et al., 2016).

Decentralization of data aims to address the market power that arises due to a few organizations controlling large volumes of data. Decentralized data storage might reduce the barrier to market entry and reduce switching costs for consumers. Moreover, decentralized data storage holds the promise to safeguard privacy sensitive information by design.

Privacy externalities and network effects, however, are likely to remain present in the case of a decentralized web. For many digital services and products, applications will still become more valuable when the number of users increases. These network effects might also prevent decentralized solutions to gain momentum without further regulation e.g. by stimulating the use of decentralized solutions or gradually putting restrictions in place for data storage by organizations. Governments could also consider piloting decentralized solutions to better understand opportunities and weaknesses (see Buyle et al., 2019, for an example of the Belgian government and MyData[22] for examples in Nordic countries).

4.3.2 Processing of data: mandatory data unbundling

In chapter 2, we saw that combining data is a central part of the data value chain. Digital platforms own part of their success to their ability to infer information about users by leveraging data from different sources. Condorelli and Padilla (2020) describe how data abundant firms can in fact obtain a dominant position in multiple platform markets by tying privacy policies, i.e. setting up privacy policies in such a way that they allow for combining data from different products or services. There are several market failures that are likely exacerbated by privacy policy tying. First, market power of data rich firms rises, resulting in higher profits and less consumer welfare. Second, tying privacy policies widens the information gap between consumers and

---

[21] T. Berners-Lee, 2018, One Small Step for the Web… (link)
[22] MyData, A Nordic Model for human-centered personal data management and processing (link)

firms as consumers loose track on how their data is used. Third, combining data from different products and services increases the likelihood of (unwanted) privacy externalities to occur.

Condorelli and Padilla (2020) suggest mandatory data unbundling within dominant firms as a possible remedy against the negative effects of privacy policy tying. Unbundling forces firms to use data solely for the product or service through which the data was obtained. This reduced scope of application makes it less attractive for firms to (over)collect data. Mandatory unbundling can also have negative consequences. For instance, rich datasets based on multiple data sources also have the potential to better match supply and demand for products and services. Creating a ban on bundling data might thus jeopardize allocative efficiency.

Mandatory data unbundling seems unlikely to materialize without new legislation (see also the box on recently proposed EU regulation). In 2019, the Bundeskartellamt (the German antitrust authority) instructed Facebook to decouple personal data obtained from different products[23]. According to the antitrust authority 'the extent to which Facebook collects, merges and uses data in user accounts constitutes an abuse of a dominant position'. Facebook however appealed this decision in court. In June 2020, the Bundesgerichtshof (the German Federal Court of Justice) concluded that Facebook abused its position by depriving users of Facebook of any choice and that the Bundeskartellamt was right to impose a cease and desist order.[24] This decision is not the last word in the legal trajectory. Upon Facebook's appeal, the Oberlandesgericht Düsseldorf (The Higher State Court in Düsseldorf) decided in March 2021 to request judicial advice from the European Court of Justice.[25] needs to decide on Facebooks appeal against the Bundeskartellamt's decision. In California, meanwhile, there is a possibility that new legislation might explicitly restrict advertising based on data from different sources[26].

4.3.3 Valorization of data: opting-out of AdTech

Using data to target consumers with personalized advertisements has become one of the major business models in the data economy. Both Google and Facebook generate most of their revenue and profit via advertisement technology (AdTech). Although highly lucrative, there is growing concern about privacy infringements inherent to this business model (see for example Zuboff (2019)).

Analogous to do-not-call-registries in telemarketing it could be worthwhile to insist digital platforms to develop do-not-track alternatives without loss of functionality. Such an option could empower consumers beyond the current informed consent approach. First, a do-not-track option would need to be implemented by all platforms at once. Second, a do-not-track option is likely easier to understand compared to the current practice of long privacy agreements. A do-not-track policy has the potential to reduce the information asymmetry between consumers and firms.

Interestingly, the call for do-no-track options in the data economy seems to experience a revival after voluntary technological implementation were considered unsuccessful.[27] Tene and Polenetsky (2011) point out that before deciding to enforce do-not-track like policies, policy makers will first have to decide whether the information-for-value business model is positive or negative from a societal point of view. There is a trade-off to be taken into account between economic opportunities and safeguarding privacy.

[23] See this press release on the website of the Bundeskartellamt
[24] See the decision of the Bundesgerichtshof: link.
[25] See the press release from the Oberlandesgericht Düsseldorf.
[26] See this blogpost.
[27] See this Wired article.

Budak et al. (2014) studied the impact of targeted advertising on the internet economy empirically and conclude that do-not-track policies 'would impact, but not fundamentally fracture, the Internet economy.' Even when consumers would value privacy such that from a societal point of view do-not-track policies are legitimate, it remains to be seen whether a do-not-track policy becomes easily implementable[28]. In contrast to phone numbers for example there is no one-to-one mapping between individuals and IP addresses or other digital fingerprinting techniques. Do-not-track would thus require a different setup compared to the do-not-call registries. Options could include browser features or prevention of cross-app tracking via operating system controls.

[28] See this blogpost

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar, forthcoming, Too much data: Prices and inefficiencies in data markets, American Economic Journal: Microeconomics.

Acemoglu, D. and P. Restrepo, 2020, The wrong kind of AI? Artificial intelligence and the future of labour demand, *Cambridge Journal of Regions, Economy and Society*, vol. 13, pp.25-35.

Acquisti, A., L. Brandimarte, and G. Loewenstein, G., 2015, Privacy and human behavior in the age of information, *Science*, vol. 347, no. 6221, pag. 509-514.

Aghion, Philippe and Peter Howitt, 1992, A Model of Growth through Creative Destruction, *Econometrica*, vol. 60, no. 2, pp. 323–351.

Azar, J., E. Huet-Vaughn, I. Marinescu, B. Taska and T. Von Wachter, T., 2019, Minimum wage employment effects and labor market concentration, NBER working paper 26101.

Barrett, S., 2016, Coordination vs. voluntarism and enforcement in sustaining international environmental cooperation, *Proceedings of the National Academy of Sciences*, vol. 113, nr. 51, pp. 14515-14522.

Benmelech, E., N.K. Bergman en H. Kim, 2020, Strong employers and weak employees: How does employer concentration affect wages?, *Journal of Human Resources*, 0119-10007R1.

Bergemann, D., A. Bonatti, A. and T. Gan, 2020, The economics of social data, *arXiv:2004.03107*.

Bergemann, D. and A. Bonatti, 2019, The Economics of Social Data: An Introduction, Cowles Foundation Discussion Paper No. 2171R.

Bijlsma, M., B. Straathof, and G. Zwart, 2014, Kiezen voor privacy: hoe de markt voor persoonsgegevens beter kan, CPB policy brief

Blanchard, O. J. and L.H. Summers, 1986, Hysteresis and the European unemployment problem, NBER macroeconomics annual, 1, pp. 15-78.

Buchanan, James M., 1965, An Economic Theory of Clubs, *Economica*, vol. 32, no. 125, pp. 1-14.

Budak, C., S. Goel, J. M. Rao, and G. Zervas, 2014. Do-not-track and the economics of third-party advertising. Boston University, School of Management Research Paper, 2505643.

Buyle, R., R. Taelman, K. Mostaert, G. Joris, E. Mannens, R. Verborgh, T. and Berners-Lee, 2019, Streamlining governmental processes by putting citizens in control of their personal data, *International Conference on Electronic Governance and Open Society: Challenges in Eurasia*, pp. 346-359.

Bygrave L. A., D. W. Schartum, 2009, Consent, proportionality and collective power. In: Reinventing data protection?, pp. 157-173.

Card, D., 1996, The effect of unions on the structure of wages: A longitudinal analysis, *Econometrica*, pp. 957-979.

Carrière-Swallow, M.Y. en M.V. Haksar, 2019, The economics and implications of data: an integrated perspective, International Monetary Fund (IMF).

Cassell, J. and A. Young, 2002, Why we should not seek individual informed consent for participation in health services research, *Journal of medical ethics*, vol. 28, no. 5, pp.313-317.

Chen, S. C. en C. J. Webster, 2005, Homeowners associations, collective action and the costs of private governance, *Housing studies*, vol. 20, nr. 2, pp. 205-220.

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

Choi, J. P., D. S. Jeon, and B. C. Kim, 2019, Privacy and personal data collection with information externalities, *Journal of Public Economics*, vol. 173, pp. 113–124.

Coase, R. H., 1960, The problem of social cost. In Classic papers in natural resource economics, pp. 87-137, Palgrave Macmillan, London.

Condorelli, D. and J. Padilla, 2020, Harnessing platform envelopment in the digital world, *Journal of Competition Law & Economics*, vol. 16, nr. 2, pp. 143-187.

Coyle, D., S. Diepeveen, J. Wdowin, L. Kay and J. Tennison, 2020, *The value of data – policy implications report*, Bennet Institute for Public Policy, Cambridge and Open Data Institute.

Eisenmann, T., G. Parker and M. Van Alstyne, 2011, Platform envelopment, *Strategic Management Journal*, vol. 32, no. 12, pp.1270-1285.

Fairfield, J. A. T. and C. Engel, 2015, Privacy as a Public Good, *Duke Law Journal*, vol. 65, pp. 385-457.

Furman, J., 2019, Unlocking Digital Competition, Report of the Digital Competition Expert Panel, H.M. Treasury, United Kingdom.

Furman, J.L. and S. Stern, 2011, Climbing atop the shoulders of giants: The impact of institutions on cumulative research, *American Economic Review*, vol. 101, no. 5, pp. 1933-63.

Goldfarb, A., and C. Tucker, 2019, Digital Economics, *Journal of Economic Literature*, vol. 57, no. 1, pp. 3-43

Hilbert, M. and P. López, 2011, The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, vol. 332, no. 6025, pp.60-65.

Hill, R., C. Stein and H. Williams, 2020, Internalizing Externalities: Designing Effective Data Policies, *AEA Papers and Proceedings*, vol. 110.

Jones, C. I. en C. Tonetti, 2020, Nonrivalry and the economics of data, *American Economic Review*, vol. 110, no. 9, pp. 2819-2858.

Kirpalani, R. and T. Philippon, 2020, Data Sharing and Market Power with Two-Sided Platforms, NBER working paper 28023

Koutroumpis, P., Leiponen, A. and Thomas, L.D., 2020. Markets for data. *Industrial and Corporate Change*, vol. 29, no. 3, pp.645-660.

Loomis J., S. Koontz, H. Miller, L. Richardson, 2015, Valuing geospatial information: Using the contingent valuation method to estimate the economic benefits of Landsat satellite imagery, *Photogrammetric Engineering & Remote Sensing*, vol 81, pp. 647-56.

MacCarthy, M, 2010, New Directions in Privacy: Disclosure, Unfairness and Externalities, *I/S: A Journal of Law and Policy for the Information Society*, vol. 6, pp. 425-512.

Mansour, E., A. V. Sambra, S. Hawke, M. Zereba, S. Capadisli, A. Ghanem, A. Aboulnaga, and T. Berners-Lee, 2016, A demonstration of the solid platform for social web applications, In *Proceedings of the 25th International Conference Companion on World Wide Web*, pp. 223-226.

Mantelero, A., 2016, Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection, *Computer law & security review*, vol. 32, no. 2, pp.238-255.

Nagaraj, A., 2021, The Private Impact of Public Data: Landsat Satellite Maps Increased Gold Discoveries and Encouraged Entry, *Management Science.*

Obar, J.A. en A. Oeldorf-Hirsch, 2020, The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services, *Information, Communication & Society*, vol. 23, no. 1, pp. 128-147.

O'brien, O., J. Cheshire and M. Batty, 2014. Mining bicycle sharing data for generating insights into sustainable transport systems, *Journal of Transport Geography*, vol. 34, pp. 262-273.

OECD, 2015, Data-Driven Innovation: Big Data for Growth and Well-Being, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264229358-en.

OECD, 2020, A roadmap toward a common framework for measuring the digital economy, OECD publishing, Paris.

Ostrom, E., 1990, Governing the commons: The evolution of institutions for collective action, Cambridge university press.

Pigou, A. C. (1920). The Economics of Welfare. London: Macmillan.

Posner, E. A. and E. Glen Weyl, 2018, Radical markets – uprooting capitalism and democracy for a just society, Princeton University Press.

Reimers, I. and B.R. Shiller, 2019, The impacts of telematics on competition and consumer behavior in insurance, *Journal of Law and Economics*, vol. 62, no. 4, pp. 613-632.

Romer, P.M., 1990, Endogenous technological change, *Journal of Political Economy*, vol. 98, no. 5, pp.S71-S102.

Smith, N.C., D. G. Goldstein, and E. J. Johnson, 2013, Choice without awareness: Ethical and policy implications of defaults, *Journal of Public Policy & Marketing*, vol. 32, no. 2, pp. 159-172.

Statistics Canada, 2019, Measuring investment in data, databases and data science: Conceptual framework, *Series: Latest Developments in the Canadian Economic Accounts*. Retrieved July 8th 2020 from https://www150.statcan.gc.ca/n1/en/catalogue/13-605-X201900100008.

Susser, D., 2019, Notice After Notice-and-Consent: Why Privacy Disclosures Are Valuable Even If Consent Frameworks Aren't, *Journal of Information Policy*, vol. 9, pp.148-173.

Swedish National Board of Trade, 2014, *No Transfer, No Production - a Report on Cross-border Data Transfers, Global Value Chains, and the Production of Goods*.

Tene, O. and J. Polenetsky, 2012, To track or do not track: advancing transparency and individual control in online behavioral advertising, *Minn. JL Sci. & Tech.*, vol. 13, p. 281.

Wachter, S. and B. Mittelstadt, 2019, A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI, *Colum. Bus. L. Rev.*, p. 494.

Wdowin, J. & Diepeveen, S. (2020). *The value of data- accompanying literature review*. Bennet Institute for Public Policy, Cambridge and Open Data Institute.

Zhu, Z., M. A Wulder, D. P. Roy, C. E. Woodcock, M. C. Hansen, V. C. Radeloff, S. P. Healey, C. Schaaf, P. Hostert, P. Strobl, and J. F. Pekel, 2019, Benefits of the free and open Landsat data policy, *Remote Sensing of Environment*, vol. 224, pp. 382-385.

Zuboff, S., 2015, Big other: surveillance capitalism and the prospects of an information civilization, *Journal of Information Technology*, vol. 30, no. 1, pp. 75-89.

Zuboff, S., 2019, *The age of surveillance capitalism*, Profile books Ltd.