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.

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.