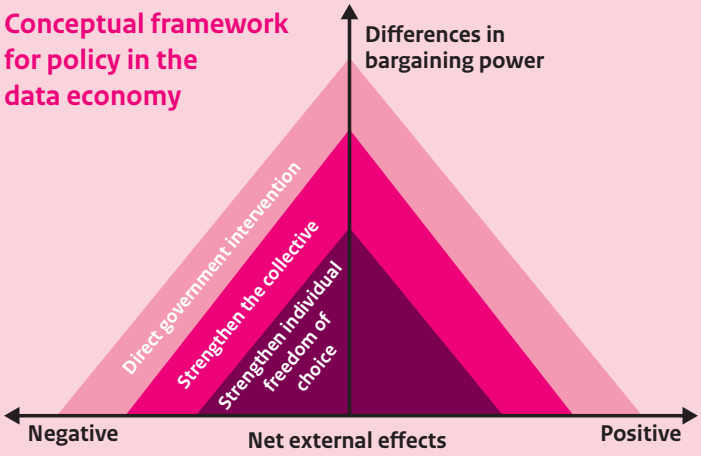
CPB Netherlands Bureau for Economic Policy Analysis

Brave new data Policy pathways for the data economy in an imperfect world



Data are playing an increasingly important role in our economy. Two bottlenecks prevent optimal functioning of the data economy: external effects and unequal balance of power between data suppliers and data processors. External effects can be positive or negative and lead to too much or not enough data being shared. The consequence of an unequal balance of power is that data processors determine the conditions of the data transaction.

We introduce a conceptual framework that structures various types of solutions for dealing with these bottlenecks.

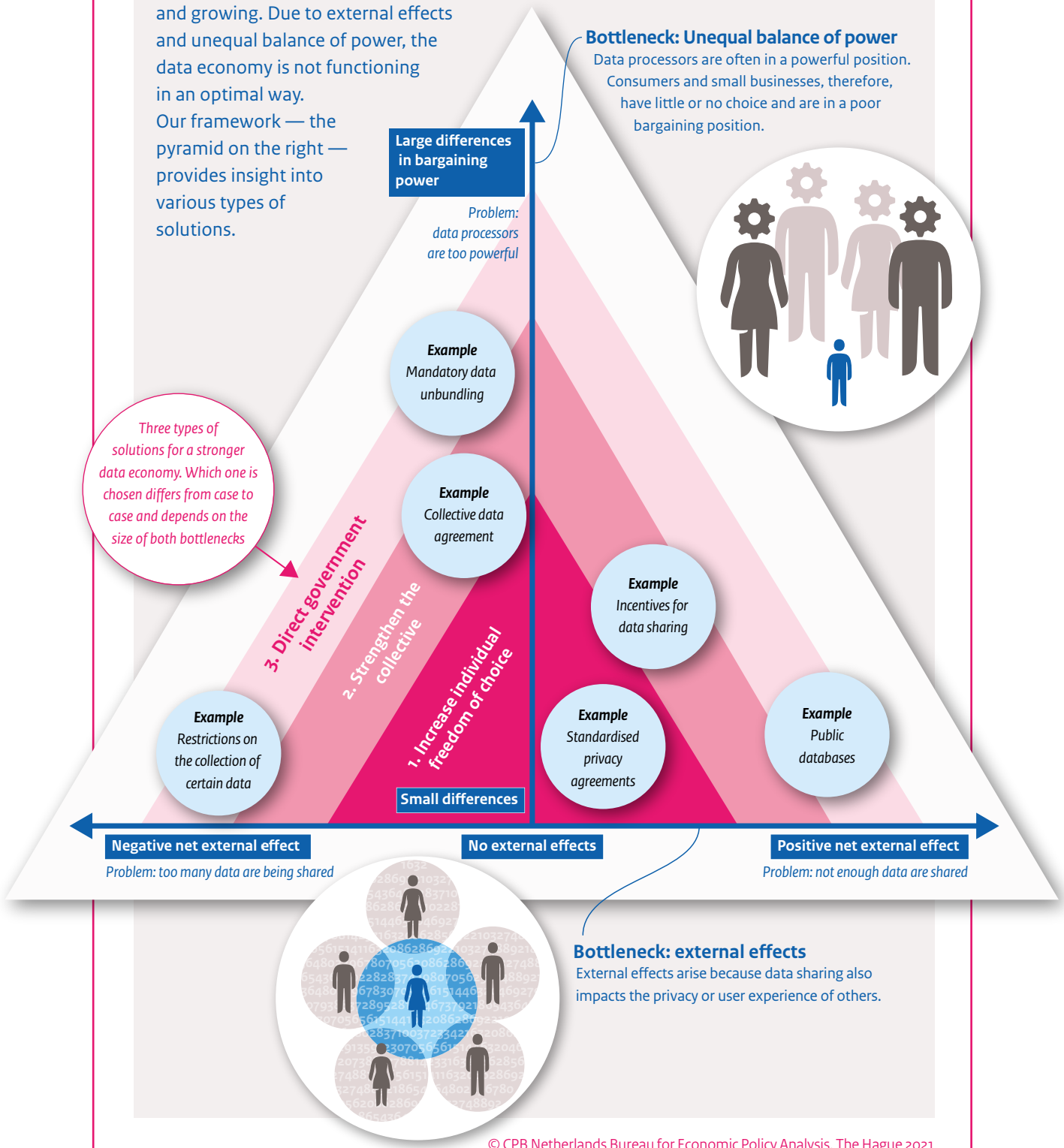
CPB Policy Brief

Ramy El-Dardiry, Bastiaan Overvest, Milena Dinkova, Rob Aalbers

July 2021

Conceptual framework for data policy

The role of data in the economy is large and growing. Due to external effects and unequal balance of power, the data economy is not functioning in an optimal way.
Our framework — the pyramid on the right — provides insight into various types of solutions.



© CPB Netherlands Bureau for Economic Policy Analysis, The Hague 2021

Summary

Data are playing an increasingly important role in our economy. Companies use data to improve and innovate their products and services. Medical data, for instance, are used to train algorithms in detecting tumours, and user data for social media platforms are analysed to develop targeted advertising. Data are also used by consumers to help them make better-informed purchasing decisions, for example through reviews or ratings.

The way data are shared often determines who are able to use them. Unlike more traditional production factors, such as labour, or consumer products and other goods, data are non-rivalrous: their use by one person does not come at the expense of simultaneous use by someone else. Nevertheless, there may be reasons for excluding others from having access to certain data. For example, because the data contain sensitive information or because exclusive access is profitable. Whether data can be used by multiple organisations or individuals, therefore, often depends on the willingness to share data.

The decision rights to share data predominantly rest with individual people or organisations. The General Data Protection Regulation (GDPR) and the Dutch vision on data sharing between companies both work from the same starting point of individual control. The GDPR gives individual rights to people, such as the right to data portability, and it obliges companies to ask each person for permission to use their personal data. The official view of the Dutch Cabinet is that data sharing between companies is preferably voluntary — although obligations are not ruled out.

Others may profit or suffer from data sharing without paying or receiving a financial compensation— in those cases where external effects occur. For example, people sharing certain data may have unintended consequences for the privacy of others. This may concern, for example, a photograph with several people that is posted on social media, or DNA data that reveal information about family members. There are also cases where data sharing leads to profits for others. For example, the availability of open data enables start-ups to innovate faster and cheaper.

Unequal bargaining power often forms a second bottleneck when data are shared. Data suppliers — usually consumers — have less influence than data processors (e.g., companies) when it comes to the final data transaction, because the latter often have market power and the interests of individual consumers in the bargaining outcome are small.

These bottlenecks lead to inefficient data sharing: sometimes too much is being shared, sometimes too little. To understand whether an increase in data sharing would have either a positive or a negative net result, an analysis is required of the direct benefits for users as well as the positive and negative impact on others. How such an analysis balances out is partially determined by the type of data. For example, privacy considerations do not always play a role in industrial data, where sharing more information is often socially desirable. And an example of too much data sharing could be the widely accessible social media profiles that are being used to influence elections.

Policies that place data sharing decisions in the hands of individual parties are not suitable for solving this efficiency problem. Individual consumers or companies insufficiently consider the possible impact of data sharing on others, nor do they have sufficient bargaining power through which to make a real choice. Current data policies, which focus particularly on individual responsibility, are therefore inadequate. Standard policy options for internalising external effects, such as pricing or assigning property rights, are not very effective where data are concerned. For example, taxing individual data transactions is complicated, and assigning property rights is not easy when data relate to multiple parties. Consumers are also unlikely to trade their data exclusively with a single party.

If the external effects or the differences in bargaining power are substantial, we see two possible solutions for a more efficient organisation of the data economy: shifting the power of choice to collectives or direct government intervention. These solutions are not a panacea, because whether government authorities always have the necessary information at hand is debatable, and because the effectiveness of certain solutions is still uncertain. Nevertheless, given the considerable negative consequences of doing nothing, implementing policies can in some cases be worthwhile.

1 Introduction

Of course, there will be challenges. But [...] the scare stories you read today about privacy concerns will just seem quaint and old-fashioned. – Hal Varian, 2013

Under the regime of surveillance capitalism [...] scientists are not recruited to solve world hunger or eliminate carbon-based fuels. Instead, their genius is meant to storm the gates of human experience, transforming it into data and translating it into a new market colossus that creates wealth by predicting, influencing, and controlling human behaviour. – Shoshana Zuboff, 2019

Digital products and services are changing our lives. Data are at the heart of this change. The two quotes above illustrate the lively debate on data use. On the one hand, data-driven applications offer economic and social opportunities. The European data strategy[1] is aimed at utilising these opportunities. Examples include personalised education and healthcare via digital means or the ability to offer them remotely (e.g., Miller et al., 2019). The possibilities of automation are increasing as algorithms are being trained to deal with large data sets. The self-driving car is an example of this phenomenon.

However, there are also growing concerns about the collection and use of data. Personal autonomy is being threatened by algorithms that are increasingly able to predict and influence our behaviour. Moreover, a handful of tech companies control most of the data. The extent to which our economy and democracy will benefit from digitalisation is still uncertain.

In the Netherlands, the coronavirus crisis revealed different views on data use. The suggestion that an app could be used for tracing potentially infected people sparked unrest. Scientists sent an urgent letter[2] to the Dutch Cabinet, calling on them to be careful when dealing with fundamental freedoms. Society is apparently still searching for ways to balance the use of data with fundamental rights and common interests.

This policy brief explores policy options for data regulation that balances common interests and individual rights. By looking at collective instruments for data sharing alongside policies aimed at individual consumers and businesses, a broader range of policy options emerges. The main question is whether, given a specific use case, it would be desirable, from a social perspective, to share either more data or fewer data compared to the status quo. The analyses and recommendations in this report are the result of a literature study and a series of interviews with scientists in the Netherlands and abroad. A more detailed discussion of the literature can be found in CPB's background document 'Policy Options for the Data Economy – a Literature Review'.

This policy brief provides an overview of possible solutions rather than answers. Chapter 4 discusses which solutions seem most promising, given the circumstances. It should be noted that there is still much uncertainty about feasibility. Careful experimentation and research will bring more clarity, over time. Some policy options also require new regulatory and institutional

[1] European Commission, *A European strategy for data*, 19 February 2020 (link)
[2] Link

frameworks. Regulation whereby, for example, the control over data is transferred from individuals to collectives requires adjusting the General Data Protection Regulation (GDPR).

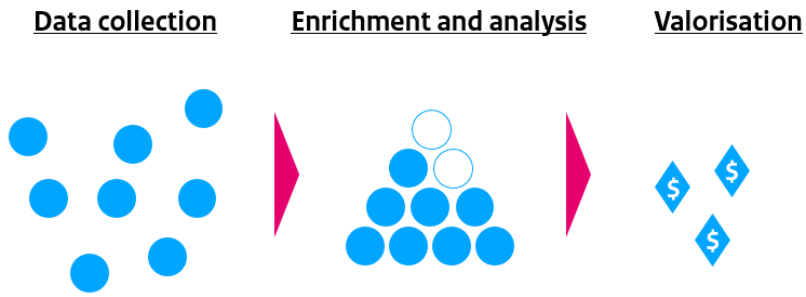
2 Characteristics of the data economy

The volume of data is increasing rapidly. Worldwide, each and every day, around 300 billion e-mails are being sent, 400 million photographs are posted on social media and 5 billion searches are performed.[3] There are billions of devices that are collecting and exchanging data.

The ability to analyse these growing quantities of data is also increasing. For example, the degree to which texts, images and sounds can be digitally decomposed and made useful is growing, due to more computing power and developments in the field of data science.

Data are used in product manufacturing and service provision and have thereby become a factor of production. Examples are traffic data feeding an algorithm for self-driving cars, internet data enabling personalised advertisements, or data used in optimising business processes. First step in the 'data value chain' (Figure 1), is data collection and storage. Next, data are translated into information by combining data and analyses. Information is ultimately converted into value through products and services. The term data economy refers to that part of the economy that is affected by activities within the data value chain[4].

Figure 1 The data value chain



In the data economy, various types of data emerge. Data often originate from administrative processes and transactions. Examples are municipal registrations and photographs posted on social media. However, data can also be created via observations or by deriving patterns from large quantities of data. The data that are released vary in the degree to which they are structured and in personal sensitivity.

[3] Source: Raconteur (link).
[4] This is in line with the definition as used by the European Commission (link): 'The data economy measures the overall impacts of the data market – i.e. the marketplace where digital data is exchanged as products or services derived from raw data – 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.'

Data are 'non-rivalrous'. This means that data may be used by one consumer without preventing simultaneous use by others (Carrière-Swallow and Haksar, 2019). Data, thus, are not like traditional private goods, such as food, furniture and clothing.

Often, data first have to be shared before they can be used. Data differ in the extent to which they are accessible to third parties. Because of these various levels of accessibility, data cannot be directly compared to traditional public goods such as clean air or dykes. Sometimes, there are legal obligations, such as in the case of income statements, which ensure that data are publicly accessible. In other cases, data are so easy to obtain that they are impossible to fence off — such as weather data, for example. Usually, however, a choice has to be made about whether or not to make data available to others. Such choices are made, for example, when consumers enter into transactions with companies or when companies decide to exchange data.

Current policy leaves the decision about data sharing primarily in the hands of individual consumers and companies. The GDPR states that personal data may be processed when consumers give their consent.[5] The data processing party is obliged to sufficiently inform the consumer prior to such processing about, amongst other things, the necessity of them accessing those data. In their data sharing vision, the Dutch Cabinet discusses the economic opportunities of data sharing between companies and emphasises that such sharing should preferably be voluntary.

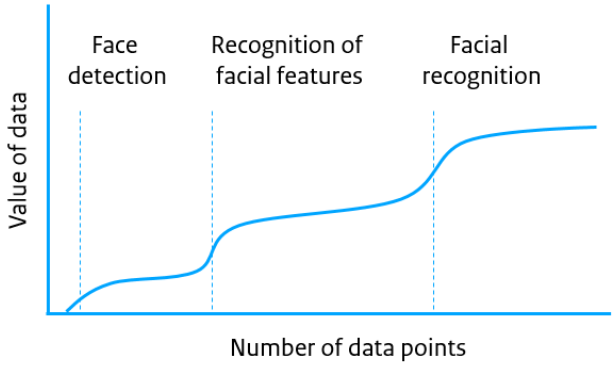
The value of data to the economy also depends on how user rights are being distributed. Data about a certain machine are valuable for the company that owns the machine, but can also be valuable for the machine manufacturer, who could use it to provide additional services or to improve the machine. Data from millions of social media users or streaming services help to target ads to specific audiences or offer personalised media. Data sharing can also serve public interest; for instance, road user data can be useful for infrastructure managers or public transport companies.

The use of data is showing diminishing returns. Data about a single user have little value for data processors unless that data enable specific user applications. Data only become valuable when companies are able to collect large amounts which can then be used in production processes. As with other production factors, data used within a certain application eventually also show diminishing returns. Any additional data points will then lead to increasingly smaller performance improvements.

However, there are also times when more data enable new applications leading to an increase in returns. Consider for example algorithms that are intended to detect human faces in photographs. Above a certain amount of data, this algorithm hardly becomes more accurate. Nevertheless, it can be interesting to collect more data in the form of photographs, because the additional data enable other, more complex applications. Photographs could, for example, be used to train algorithms to distinguish facial features and ultimately identify people. The value of data then increases by leaps and bounds, see Figure 2.

[5] See Article 6 of the GDPR.

Figure 2 Illustration of the impact of additional data on their value



Source: adapted from Posner and Weyl (2018)

What are the bottlenecks? The data characteristics as described above are not unique; being non-rival and having limited excludability applies also to other things. Air or dykes, for example, are also non-rivalrous. Problems arise or opportunities are missed in cases of market failure. The following chapter discusses specific market failures in the data economy.

3 Bottlenecks in the data economy

Two bottlenecks are preventing optimal functioning of the data economy: externalities and unequal balance of power between data suppliers (i.e., often consumers or smaller companies) and data processing organisations. External effects occur when individual choices also affect the privacy or user experience of others, and because organisations are able to use data that is released from third-party data transactions. This chapter further discusses these bottlenecks. Current data policies do not yet have a good answer to these bottlenecks. This suggests that welfare can increase through policy adjustments.

3.1 External effects

External effects occur when someone's personal data also affects the privacy or user experience of others. People's personal information is often intertwined. Examples are photographs that show multiple people, recordings of conversations between various people, and DNA that contains revealing information about relatives. These are all examples of data that transcend the individual. In addition, personal data can be used to make predictions about other people. This last situation occurs when data processing companies use large quantities of data to create profiles of people on which they have no data (MacCarthy, 2010). Companies are able to use the data on a particular group of people to derive widely varying types of information about other people, such as political preferences, sexual orientation or ethnic background.[6] An important consequence of this correlated information is that one person's choice to share data with a company ('the data transaction') has an impact on the privacy or user experience of others who are not involved in the data transaction

[6] See this article on *Ars technica*.

(Acemoglu et al., 2019). For individual people, such external effects can be either negative or positive. For example, influencers on social media may welcome their image being shared by others on the internet, whereas parents may not appreciate other people posting a photograph that also includes their child. If medical scans are used in training diagnostic algorithms that improve the quality of diagnosis over time, patients in the future will benefit. The magnitude of the external effect also depends on the type of data and the way in which these data are used.

External effects may also occur when organisations enter into data transactions. When data of a certain company, public or semi-public organisation become accessible to others, they could be used in novel ways. This may be the case in the healthcare sector, for example; companies specialising in artificial intelligence can train algorithms based on scans or ultrasound images to improve the accuracy of diagnostic tools. And medical scientists can use data from multiple healthcare institutions to determine whether a treatment is effective or what side effects occur. External effects can also be negative when data are used by organisations. For example, when a data processing company invests little in cyber security, it may not fully take into account the possible impact that a data leak can have on others (CPB, 2016).

Because of these external effects, data are being shared inefficiently. As Chapter 2 describes, the starting point of current data policy is that consumers and companies should determine for themselves whether or not they will share their data. Consumers are free to share their personal data with any data processing organisation, also if their data indirectly contain information about other people. And companies may, for fear of competition, often not be inclined to share their data or make them publicly accessible (Jones and Tonetti, 2020). The potential social value of data is then insufficiently utilised. From society's point of view, external effects may lead to either too much data being shared or too little, depending on the type of data and how they are used.

There are no obvious solutions to this problem. The usual solutions for internalising external effects are not very effective where data are concerned. A traditional solution is to define property rights. For example, the emission of harmful substances such as CO_2 or nitrogen oxides is reduced in various countries via a system of tradable emission rights (e.g., the EU ETS). Where data are concerned, assigning exclusive property rights is not always as easy. For example, in the case of data resulting from financial transactions; both sides of the transaction automatically have these data in their possession. In addition, consumers are unlikely to trade personal data exclusively with a single party. After all, consumers can share data with others as much as they like, without these data losing their functionality.

Occurrence and magnitude of external effects also depend on the way data are used at a later point in time. Consider user data on social media platforms. Platforms can utilise these data in many different ways. They can, for instance, change their own user interface or offer personalised advertisements to specific users. In both cases, there are external effects — the product on offer changes per user, but these users may all differ in their opinion of these external effects. Platforms also have the option of not immediately using these data, which means that external effects are absent, at least temporarily.

The pricing of data transactions with external effects is also a complicated matter. Pricing is the second traditional solution to internalise external effects. It could, for example, take the form of a subsidy on transactions with positive external effects or an excise duty on transactions with a

negative external impact. This is also a solution that is difficult to implement, in practice, because the size of the external effect depends on the type of data, the individual privacy preferences of consumers, and the type of application (e.g., whether data are used for scientific research or for pricing insurance). Moreover, monitoring all data transactions and the way data are used is practically impossible.

3.2 Unequal balance of power

In addition to external effects, an unequal balance of power is a second type of bottleneck in the data economy. This section discusses how such inequality comes about.

Bargaining power is often unevenly distributed within the data economy. People are consciously, and unconsciously, providing data, e.g. when they consume a digital service or when they are monitored digitally at work, but often have limited control over or insight into how their data are used by data processors. This unequal balance of power has several causes: market power, information asymmetry and behavioural biases. As a result, mainly the data processors determine, within the legal framework, which personal data are collected and how they are used, rather than the party supplying the data.

In certain sectors, consumer choices are limited due to market power. Some platform services are in a dominant position, making it difficult for consumers to 'vote with their feet', so to speak, with respect to the platform's data policy. For example, if all your friends are on Facebook or WhatsApp, you will be more inclined to accept the platform's privacy terms than if you have more platforms to choose from. This problem also occurs in other sectors: government agencies often have a monopoly and can unilaterally adjust data conditions within legal limits.

Moreover, information asymmetry and behavioural biases can worsen the bargaining position of consumers. Consumers cannot see what companies are doing with their data, nor can they read all of the data terms of use or oversee the consequences. Companies are able to exploit their strong informational position by manipulating the preferences of consumers and enticing them to buy products or share more data (Acquisti et al., 2015). Examples can be found at booking websites and online retailers who deliberately suggest scarcity in order to entice consumers to buy products.

An unequal balance of power may also occur between companies. The data of certain companies may also be of use to other companies in the development of new services or products. Such data may represent essential input in product development. If there is only one company with these essential data, the receiving companies have little bargaining power. For example, after the Cambridge Analytica case, Facebook changed the terms of use of its API (application programming interface). Stricter conditions were partly necessary to reduce the risk of abuse. At the same time, smaller companies and developers were put under pressure by these new and stricter terms for the

use of essential data, to the point where they got into trouble. Sometimes, platforms deny access to data to emerging competitors to exclude them from competition[7].

European policy is aimed at reducing the unequal balance of power. The GDPR gives individual rights to each data supplier (i.e., consumer), but because of the large number of data suppliers and the small interests per individual data supplier, their negotiating power with the data processor is unbalanced. The existing policy (perhaps with the exception of antitrust regulation) does not affect the balance of power between companies. The European Commission, therefore, is developing new regulation to equalise this balance of power, including the Data Governance Act, the Digital Services Act and the Digital Markets Act. For example, large platforms will be obliged, under certain conditions, to allow third parties to develop services that make use of their platform (interoperability). This proposed legislation also provides for the possibility of setting up organisations that pool the interests of small businesses and consumers.

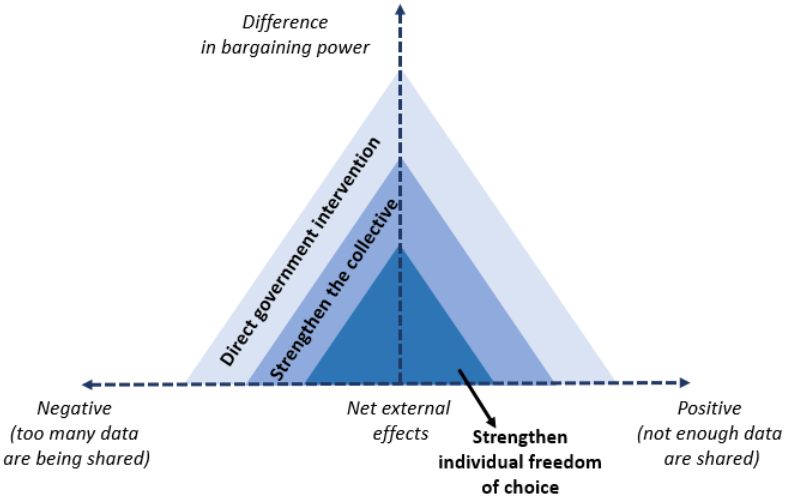
[7] Twitter, for example, denied access to the data analysis company DataSift after it had taken over a comparable business itself. See this message for a discussion on multiple relevant cases.

4 Policy pathways

4.1 Conceptual framework for policy options in the data economy

This chapter introduces a conceptual framework that structures various policy solutions. The framework, shown in Figure 3, consists of three layers. The optimal way for governments to intervene depends on both the size and nature of the net external effects. Within the innermost layer of the pyramid, external effects and the difference in bargaining power are small. In such situations, citizens and companies can weigh up the pros and cons of data sharing themselves.[8]

Figure 3: Conceptual framework for policy in the data economy



Collective action comes into play when the external effects or the differences in bargaining power are substantial. The idea here is that groups of consumers jointly enter into negotiations with another party. The traditional set of instruments favoured by economists, such as changing consumer and company behaviour via monetary incentives, will only partially be effective here. This situation is represented by the middle layer of the pyramid. Policy options for the middle layer are discussed in Section 4.3.

Direct government intervention makes sense when data-related external effects or differences in bargaining power are very large. In such cases, governments can intervene directly through restrictions on data use and by imposing obligations, or by collecting and providing access to data

[8] Although there are a number of options to further strengthen this policy, they are not discussed in this policy brief. See the background document 'Policy Options for the Data Economy – a Literature Review' for a more extensive review.

themselves, see Section 4.4. Before individual policy options are discussed, the next section first explains how the framework can be used in practice.

The analysis in this chapter is limited to the economic perspective. When and where to intervene in the data value chain is, however, not purely an economic problem. There may for instance also be ethical reasons for not collecting certain types of data, such as on ethnicity.

4.2 Applying the conceptual framework in practice

The framework helps in identifying solutions, but for specific cases it is often unclear where problems are located within the framework, a priori. Consider the external effect of privacy violations caused by data sharing. In many cases, the government has insufficient information to determine the extent of this external effect. The importance that individuals attach to privacy differs from person to person and also depends on the type of data, the data processor and the type of application. Moreover, the rationale for or necessity of government intervention partly depends on the societal context.

Insight into the nature and scope of the problem can nevertheless be helpful. The starting point for thinking about possible solutions requires a good understanding of the type of data and their use. This then enables identification of possible external effects and the consequences of differences in bargaining power. As soon as the problems for a certain type of data and application have been identified, a public debate allows for weighting the external effects and finding the right type of government intervention.

In the case of external effects, the net impact determines whether governments should stimulate or restrict data use for a given application. External effects associated with the use of a certain type of data can be both positive and negative. Take the example of DNA data used by hospitals for scientific research. For part of the population, loss of privacy does not outweigh potential benefits, while for others it does. Moreover, hospitals do not always have the right incentives to share the data they possess. Governments should weigh these external effects to decide whether more or fewer of such DNA data should be shared in such cases and what the preconditions should be.

Data regulation is a complex matter. The external effects of data sharing depend on a number of factors, such as the type of data, the owner, the purpose of the application and the people or legal entities these data refer to, which is why determining the net scope of these external effects is rather complicated. The number of cases to be investigated is infinitely large. It is neither realistic nor efficient to regulate all these cases separately. To deal with this complexity, prioritisation is required — which applications appear a priori to be the most problematic and for which applications will existing regulation be sufficient? Moreover, bottlenecks in the data economy often transcend national borders. Large data processors are often located outside the Netherlands or the European Union. Effective regulation and supervision, therefore, require international coordination or agreements.

Given this complexity, building knowledge within government institutions is indispensable for social debate on data regulation. In 2020, the Dutch House of Representatives set up a committee

to gain more control over digitalisation. Although several ministries, advisory councils, supervisory authorities and universities all provide information about data and digitalisation, for members of parliament it remains difficult to fully understand the impact of digitalisation.[9] Given that the societal impact of digitalisation is researched at many different places, there is the risk of fragmentation. Discussions on external effects and unequal bargaining power are then likely to remain underdeveloped.[10] To prevent fragmentation and build a solid knowledge base, research agendas could be better coordinated and specific knowledge institutions could be founded.

4.3 Strengthening the collective

Collective action is an opportunity for improving the data economy. External effects can be taken into account when negotiations about data sharing and use are shifted from individuals to communities. Strengthening the cooperation between data suppliers may bring external effects to light, for example in the area of privacy. Internalising external effects by creating collective action would also be advantageous for governments, as authorities themselves would not need to have insight into the exact scale of external effects.

Data suppliers who join forces increase their bargaining power when dealing with data processors. Moreover, for some forms of data collection and use, individual consent is unrealistic, *a priori*. Think, for example, of 360-degree street view photography.

One way to achieve more collective action is the formation of collective data agreements (CDAs). Under such CDAs, a group of data suppliers (e.g., consumers) would enter into an agreement with certain data processors (e.g., companies) about data transactions.[11] A group of data suppliers acting together makes it more likely that an economically optimal outcome is reached. Such a collective agreement could take various forms, such as more consumer-friendly general terms and conditions or, in a more extreme form, collective contracts. People relinquish their individual freedom of contract when signing such a contract, similar to what happens in the case of collective labour agreements. The use of such CDAs can be experimented with in places where natural groups of 'data suppliers' already exist, such as schools (e.g., parent councils), work (e.g., works councils) or sports associations (e.g., management boards or general member meetings).

CDAs are new and their practical implementation is uncertain. Since CDAs do not yet exist, possible practical obstacles to successful negotiations are still unknown. Two main challenges can be envisaged. First, the incentive for individuals to organise themselves into a collective is rather low. Data transactions are often negligible events for individual consumers. Free-rider behaviour may result in nobody wanting to invest any time in a CDA negotiation. We expect there to be substantial differences in level of concern between users, particularly with respect to large internet platforms. The second challenge consists of the fact that, for data processors, the benefits of a CDA

[9] Dutch House of Representatives, Report temporary committee 'Digitale toekomst - Naar meer parlementaire grip op digitalisering, 2020' [*digital future; towards greater parliamentary control of digitisation*] (link)
[10] Also see the recent AWTI report 'Rijk aan kennis - Naar een herwaardering van kennis en expertise in beleid en politiek' [*rich in knowledge, to a re-evaluation of knowledge and expertise in policy and politics*], 2021 (link)
[11] This idea has parallels with *community consent* in medical research. In a blog post, Jeni Tennison describes *community consent* for the data economy (link). This blog post (link) by Anouk Ruhaak discusses *collective consent*,

may be too small. These benefits largely consist of lower transaction costs that are currently induced by the informed-consent model.

In addition to these organisational challenges, CDAs also require new legal and institutional frameworks. The proposed EU Data Governance Act allows for data cooperatives. CDAs could become an extension to the GDPR. This could be an exemption under conditions of the GDPR or a legal obligation to enter into a CDA for specific cases. In addition, the incentive to enter into a CDA can be increased by changing the outside option (i.e., the situation without a CDA). New frameworks also need to describe how to monitor a CDA and what happens in cases of disagreement.

Collective action can also reduce bottlenecks in data sharing between organisations. Cooperation between companies offers an opportunity to internalise external effects in the data economy. For instance, organisations benefit collectively from sharing knowledge and data on cyber attacks and threats. Unequal bargaining positions between organisations arise, for example, in cases where data from one company are essential for a large number of other companies or when smaller organisations give their data to a large data processor, often a tech company. Also in these cases, the difference in bargaining power between organisations may be reduced through collective negotiations or, alternatively, by prohibiting discriminatory access conditions.

In cases where the net external effects of data sharing between companies are large and positive, creating extra incentives for data sharing makes sense. Incentives could consist of only allowing access to a public database to companies that also contribute their own data, or by making data sharing a precondition for subsidies or market entry. A possible disadvantage of these types of measures is that organisations have little incentive to share data with organisations who, themselves, have no data to share. As a result, potential welfare gains from data sharing might remain unexploited.

4.4 Direct government interventions

When net externalities or differences in bargaining power grow larger, direct government intervention may be needed. This is not unique to the data economy. Fundamental research, for example, has positive external effects and is, therefore, subsidised by the government. Consumer protection legislation reduces differences in bargaining power. In addition, competition authorities safeguard competition in an economy. Recently, various proposals have been made to strengthen this, partly extraterritorial, supervision in the data economy (Crémer et al., 2019 and EZK, 2019-2).

In addition to the various forms of supervision, governments can intervene more directly in the data economy in three ways: via legal obligations, restrictions, or by collecting and providing access to the data themselves. This section further explains a number of possible policy directions together with their advantages and disadvantages. Some of these options are discussed in more detail in the background document.

Within the data value chain, the government has various options to improve the balance of power by imposing obligations. For example, government authorities on either national or EU

level, could oblige platform companies to only process personal data that are locally stored in a personal data space of consumers.[12] A personal data space gives consumers more control over their data and reduces the costs that are related to switching between services. This increases the bargaining power of consumers. In the second part of the data value chain — where data are analysed and enriched — government authorities could intervene by obliging companies to separate their various data-driven services (i.e., mandatory data unbundling, see Condorelli and Padilla (2020)). It is easier for consumers to understand what happens to data when data processing is limited to one particular product or service. Mandatory data unbundling reduces information asymmetry. Differences in bargaining power between companies can be addressed by extending the already existing right to data portability to business users of platforms or by requiring services to be interoperable.

In the case of net negative externalities, measures aimed at reducing data collection make more sense. An obligation to offer a certain type of business model, such as subscriptions, reduces the incentives for companies to collect data. This option has the disadvantage that certain products and services might become available to only a small group of consumers, or that privacy becomes a luxury good. Mandatory unbundling of data, as mentioned earlier, is another way to reduce the incentive to collect as many data as possible. An advantage of measures such as unbundling and imposing a certain business model is that they are generic; the government does not need specific information on the external effects of data to apply them.[13] However, it remains uncertain to what extent these measures will reduce external effects.

Governments could impose a ban on collecting certain types of data. For example, authorities may require organisations to have a licence for the storage and analysis of DNA data. Furthermore, the government could prohibit certain data applications, while allowing the use of the same data for scientific research.[14] This form of regulation, however, can quickly become rather complicated. External effects are determined not only by the type of data but also by their application and type of data processor. This creates many combinations that potentially require separate regulations. In addition, the government does not always have the necessary information at its disposal to properly estimate the scope of external effects.

Finally, in those cases where net externalities are positive, setting up public databases is a policy option. A public database makes data accessible, ensures that insights are spread more widely and enables new applications. In this way, society would benefit from the non-rivalrous character of data. Prime examples of existing public databases are those of national statistical institutes, who make data from various sources accessible at an aggregated level for the general public and, under strict conditions, at a micro level for scientific research. New socially valuable databases are conceivable in various sectors. Examples include data released through the use of sharing platforms.

[12] This proposal is by Tim Berners-Lee, amongst others (link).
[13] In the field of cyber security, obligations can also reduce externalities. Consider, for example, requirements for unique passwords for Internet of Things devices and mandatory encryption for network traffic.

[14] An example of prohibiting an application is the US Genetic Information Nondiscrimination Act, which was passed in 2008. It protects citizens against the use of genetic information in insurance or in an employment relationship.

For example, shared mobility data provide insight to governments with respect to infrastructure bottlenecks, and allow companies to develop innovative applications[15].

A risk with public databases is the possible infringement of fundamental rights such as the right to privacy or equal treatment. A successful database can only exist through careful data management that safeguards cyber security, quality and privacy. It may be of strategic importance for the government to make these investments, for example to help solve major societal challenges.

5 In conclusion

Digitalisation is an important source of economic growth and offers opportunities for solving societal challenges. Data are essential in this process. However, the market for data is not working properly due to external effects and unequal balance of power. As a result, sometimes, too many data are being shared and, at other times, too few. The policy framework and the possible solutions described in this Policy Brief are not definitive answers, but provide possible pathways for the future. Much is still uncertain, for instance about the circumstances in which collective data agreements offer a solution or about the types of data that should be shared more often. New insights are gradually developing. The speed and impact of the digital transformation do require a solid knowledge base and agility.

[15] oneTransport is a UK example of this (link)

References

Acemoglu, D., A. Makhdoumi, A. Malekian and A. Ozdaglar, 2019. Too much data: Prices and inefficiencies in data markets, NBER working paper 26296.

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

Carrière-Swallow, M.Y. and M.V. Haksar, 2019, The economics and implications of data: an integrated perspective, Internationaal Monetair Fonds (IMF).

Condorelli, D. and J. Padilla, 2020, Harnessing platform envelopment in the digital world, *Journal of Competition Law & Economics*, vol 16, no. 2, pp. 143–187.

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

CPB, Cyber Security Risk Assessment (CSRA) for the Economy, 2016.

CPB, Cyber Security Risk Assessment (CSRA) for the Economy, 2018.

EZK, 2019, De Nederlandse visie op datadeling tussen bedrijven [*the Dutch vision on data sharing between companies*], Ministry of Economic Affairs and Climate Policy .

EZK, 2019-2, Discussienotitie: toekomstbestendigheid mededingingsbeleid in relatie tot online platforms [*discussion paper: future-proof competition policy in relation ot online platforms*], Ministry of Economic Affairs and Climate Policy.

Jones, C. I. and C. Tonetti, 2020, Nonrivalry and the economics of data, *American Economic Review*, vol. 110, no. 9, pp. 2819–58.

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.

Miller, A.C., Obermeyer, Z. and Mullainathan, S., 2019. A comparison of patient history-and ekg-based cardiac risk scores, *AMIA Summits on Translational Science Proceedings*, 2019, p.82.

Posner E.A. and E. Glen Weyl, 2018, *Radical Markets*, Princeton University Press.

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.