Aktuelle politische
Debattenbeiträge

IW-Policy Paper 21/20

Open data and data sharing

An Economic Analysis

Alevtina Krotova, Armin Mertens, Marc Scheufen

Köln, 13.10.2020

INSTITUT DER DEUTSCHEN WIRTSCHAFT

Table of contents

JEL-Classification:

L21: Business Objectives of the Firm
L86: Information and Internet Services • Computer Software
M21: Business Economics
O32: Management of Technological Innovation and R&D

Abstract

Data is an important business resource. It forms the basis for various digital technologies such as artificial intelligence or smart services. However, access to data is unequally distributed in the market. Hence, some business ideas fail due to a lack of data sources. Although many governments have recognised the importance of open data and already make administrative data available to the public on a large scale, many companies are still reluctant to share their data among other firms and competitors. As a result, the economic potential of data is far from being fully exploited. Against this background, we analyse current developments in the area of open data. We compare the characteristics of open governmental and open company data in order to define the necessary framework conditions for data sharing. Subsequently, we examine the status quo of data sharing among firms. We use a qualitative analysis of survey data of European companies to derive the sufficient conditions to strengthen data sharing. Our analysis shows that governmental data is a public good, while company data can be seen as a club or private good. Latter frequently build the core for companies' business models and hence are less suitable for data sharing. Finally, we find that promoting legal certainty and the economic impact present important policy steps for fostering data sharing.

1 Introduction

Data is an essential resource for a plethora of digital trends and technologies, such as artificial intelligence, the internet of things or smart services. Firms that hesitate to digitalise their business risk falling behind their competitors in the long run. Some firms and small companies or start-ups have innovative business ideas, but lack access to the necessary data sources. In this context, the importance of open data has become apparent in the past couple of years. For Germany, some studies estimate the economic impact from providing different sources of data openly to at least €2.5 billion (see Kuzev et al., 2016 for an overview). The European Commission focusses on this important topic in the European Data Strategy presented in February 2020, emphasizing the strength of a joint European data space, which should serve as the basis for future technological and economic development of the European business and public sector (European Commission, 2020a). This data space should contain data provided by governmental organisations as well as by European companies.

The amount of open data actively published by European countries, federal states and single municipalities has been increasing steadily in the last couple of years (OECD, 2018). This data can be accessed freely by all interested parties. However, simply publishing data online is not enough to generate surplus value for companies accessing the data (Kuzev et al., 2016). The surplus value is created by companies via processing, analysing and combining several data sources into valuable information. This information can be used, for instance, for cost reduction, for achieving higher growth through new business models, or for the optimisation of existing business models. The range of opportunities is as heterogeneous as the companies accessing the data.

While the innovation potential of governmental data has already been recognised, vast amounts of data held by companies is still widely unused and remains hidden in internal database silos. To fill the future European data space with business data, companies in Europe should be incentivised to share their data with their partners, suppliers and customers, but also in an interdisciplinary and intersectoral ecosystem. However, the readiness to share one's own company data with other stakeholders is currently not sufficiently high (Demary et al., 2019). Governmental initiatives like a data space for the German automotive industry face strong resistance from certain companies (Delhaes, 2020).

The reluctance of economic actors to share data has been subject to several economic studies. According to a OECD report, the main barriers to data access, sharing and re-use of data are concerns about data protection and data ownership (OECD, 2019). Some companies have difficulties to identify which data can be shared and under which conditions. Another obstacle is the fact that data are often stored in organisational silos. This implies high investments connected to the provision of high-quality data. Furthermore, it is a challenge for companies to determine the market value of data. A great number of firms claim that they would only share data if they benefited from sharing and if others shared data as well (ibid.). Richter/Slowinski (2019) find that the fear of insufficient transparency and market power aggregated by some players prevent most companies from sharing their data, while emerging data sharing platforms can resolve these issues by proving the suitable framework conditions. Shen et al. (2020) identify three main

challenges for data sharing: integrity, security and usability concerns, all resulting from strict privacy requirements.

Against this background, the aim of this policy paper is to identify the incentives needed to overcome the obstacles for data sharing and to provide suitable policy recommendation. The process of data sharing rests on the principles of open data, which are common in the context of governmental data. It is therefore important to understand the characteristics of open governmental data in order to use them as a blueprint for successful implementation of data sharing in a business context. We start with a short overview on the open data movement. Subsequently, we examine the development of open data initiatives by governments. Next, we compare the characteristics of open governmental data and shared business data, highlighting decisive similarities and differences. We provide an overview of the status quo of data sharing among EU companies and focus particularly on the required incentives for increased data sharing in chapter 3. For this purpose, we analyse results from the public consultation on “Building a European Data Economy” conducted by the EU Commission in 2017. We complete our analysis with five recommendations for action by politics as well as companies.

2 The open data movement

The open data movement can be subsumed under a broader movement entitled “open knowledge”. First, we briefly sketch the general idea of the open data movement, presenting two contradicting views on sharing data. Having laid the foundations on the idea of openness, we subsequently address the open data movement in greater detail.

2.1 The idea of open data

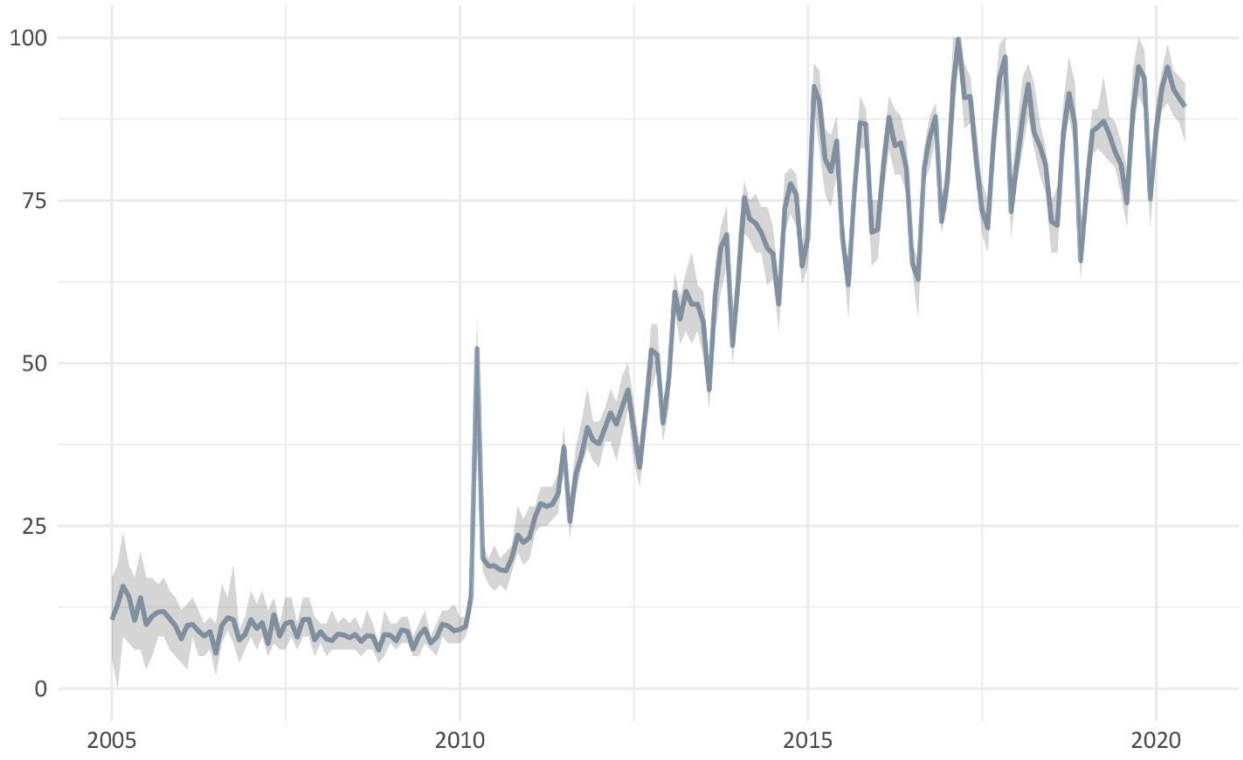
The history of open data and the open data movement eventually starts in 1957 with a focus on scientific data. In the context of the international geophysical year from July 1st, 1957 until December 31st, 1958, researchers in geophysics laid the foundations for the implementation of data sharing centres and the standardization of metadata. The first mention of the term open data stems from the 1970s, where NASA used an open data policy in a directive directed towards international collaborative partners, requiring them to adopt an open data policy similar to that of NASA and other US agencies. In particular, the directive specified not only which kind of data had to be provided for operating US-satellites, but also the data format (Yu/Robinson, 2012).

It was not until the 1990s and early 2000s that the open data movement gained momentum as a consequence of the advent of the internet and thus the technological means to share data freely via the web. Most interestingly, the movement broadened its scope of coverage from research data to governmental data. Regarding the property of data as an information good, the works of Ostrom (Ostrom, 1990, 2009) highlight the special character of the information commons. She argues that information shall be seen very similar to public goods, as information is characterized by non-rivalry in consumption, i.e. the use of information by one person does not harm the utility derived from the same use by another person. The non-rivalry of information, however, is a good fortune as it reveals the non-scarce character of information. From an economic perspective, the good should be consumed as much as possible as each consumption adds a positive surplus to social welfare. Most interestingly, the use of the information does not only lead to a non-reduction of its stock, but to an enrichment of it. The general property rights theory approach to the public good problem, in contrast, argues in favour of a restriction of the access to information, as a public good is not only characterized by non-rivalry but also non-excludability (Rusche/Scheufen, 2018). The non-excludability character of information finally induces an incentive problem, since non-excludability leads to a free riding on the investment of the creator. Hence, an intellectual property right (e.g. a patent or copyright) should foster the creation of new knowledge or information by means of an exclusive right to commercially exploit the innovation. A patent or a copyright licence can hence be seen as a temporary monopoly seeking to correct the market failure resulting from the public good character of information.

Google Trends data shows that the interest on the topic “open data” among Google users started to continuously increase since 2010 (see Figure 2-1). Since 2015, the search interest in open data has remained on a constantly high level, finding its absolute maximum (so far) in 2017. We expect the trend to persist in the future as open data is becoming a relevant topic worldwide.

Figure 2-1: Worldwide interest in open data

The blue line indicates the relative number of Google hits for the topic “open data” in a given month in relation to the maximum number of monthly hits in the time period under investigation. Considering the variance in possible values for each month caused by Google’s sampling procedure, the blue line shows the mean number of hits for 100 separate queries. The area in shaded grey illustrates the maximum and minimum values for each month.

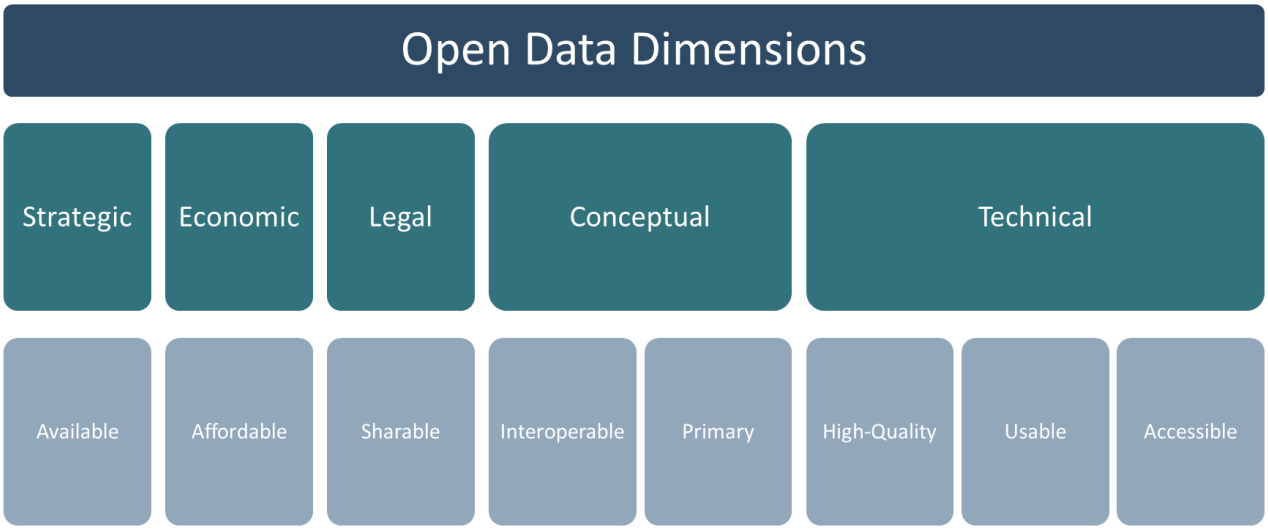


Source: Google Trends, 2020; own representation

In recent years, the definition and common understanding of the term open data has been shaped by academics and non-profit organisations having a close connection to the open data movement. We follow the frequently used definition by the Open Knowledge Foundation - OKF (2015) and define open data as follows: “Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness). Open data and content can be freely used, modified, and shared by anyone for any purpose” (OKF, 2015). The OKF is a global, non-profit network founded in 2004 that promotes and shares information at no charge, including both content and data (OKF, 2020). Other noteworthy contributions to the understanding of open data are provided by another non-profit organisation that advocates open governmental data – the Sunlight Foundation. Its ten principles of open data published first in 2010 still represent the standards for the characterisation and classification of open governmental data (Sunlight Foundation, 2017). Both contributions to the open data definition are state of the art and commonly cited by other organisations and researchers (see e.g. Tammisto/Lindman, 2012; Kalampokis et al., 2013; Wewer, 2020). In scientific research, Janssen et al. (2012) provide frequently cited structural research on benefits and barriers of open data as well as on the framework for comparing open data policies. The contribution by Jetzek (2016) in comparison, sheds new light on the dimensions of open data.

As open data is a relatively new topic in the scientific discussion, further research is needed to better understand this phenomenon and its economic, political, social, and technical effects. On the basis of the Sunlight Foundation's principles and the dimensions developed by Jetzek (2016), we identify eight common characteristics of open data: they must be available, affordable, sharable, interoperable, primary, high-quality, usable and accessible (see Figure 2-2). These eight characteristics belong to five broader dimensions of open data: strategic, economic, legal, conceptual, and technical. We assume that these characteristics build the framework of necessary conditions for data sharing.

Figure 2-2: The dimensions of open data



Source: Sunlight Foundation, 2017; Jetzek, 2016; own representation

The following paragraphs briefly describe each of the characteristics:

Available Open Data:
The access to data is non-discriminatory and not subject to any kind of limitation. This means that any person can access the data without any registration or membership being required. The platform where the data can be accessed is open and without any access restrictions.

Affordable Open Data:
The access to data is free of charge or amounts to the minimum of creating costs (e.g. marginal cost of reproduction). As data is commonly generated and collected to own purposes, the creation costs should not be passed to individuals accessing these data. Open data used by companies can positively impact business growth and indirectly benefit governments through increased tax revenues.

Sharable Open Data:
Data is published under open licences, i.e. users are free to reuse the accessed data having four freedoms: to use data for any purpose, to study and adopt the data, to redistribute and copy the data, to modify the data and share the results (Free Software Foundation Europe, 2020).

Interoperable Open Data:
Formats used to store data are commonly owned or open standard. Data is accessed by freely available formats without the need for a software license, such that it is available to a wide range of people. Common standards ensure that data can be used and exchanged unobstructed, but also empower innovations.

Primary Open Data:
Data originates from primary sources and includes original unmodified information. Moreover, information about the generation process of the given data should be as transparent as possible for users, in order to be able to verify the originality of the accessed data.

High-Quality Open Data:
The provided data is as accurate, complete, consistent, valid, and unique as possible. In addition, the data is uploaded timely and is updated regularly if necessary. Poor quality of data leads to increasing costs of usage and decreases its utility.

Usable Open Data:
Data is uploaded using machine-readable formats to ensure straightforward machine processing. Storing data in common but not machine-readable file formats like PDF or JPEG leads to unnecessary processing costs and errors.

Accessible Open Data:
Data is easily, permanently, and securely accessible and downloadable if required. The access to data is provided through a platform or another common device. The presentation of data and connected tools and services is clear and compact.

2.2 Governmental initiatives on open data

In recent years, the open data movement was essentially connected to the promotion of open governmental data. Since national and local authorities and institutions generate and commission enormous amounts of data, making these data public should promote transparency and accountability vis-à-vis its citizens (OECD, 2020). In addition, creating tangible economic impact has become a major objective of the open data movement (Huyer/van Knippenberg, 2020). The European Commission estimated an open data market size of €184.45 billion for the EU in 2019, accompanied by 1.09 million open data employees as well as significant efficiency gains and cost savings in several sectors (e.g. health care, public transport and energy consumption) (Huyer/van Knippenberg, 2020).

In terms of the technical implementation, most countries offer online portals where all the available open data can be accessed. The range of data provided on such portals varies considerably. For instance, it can include data on governmental budgets and spending, national laws, air and water quality, land ownership, election results, procurement as well as indicators on business and the economy. High quality open data portals host data that are openly licensed and publicly

available free of charge. In addition, the data should be frequently updated, complete, consistent and should be provided in machine-readable format (Lämmerhirt et al., 2017). In absolute numbers, the United States provide the most open data portals on national and local administrative levels (670 open data portals). Runner-ups are France (352 portals), Spain (308), the United Kingdom (141), Canada (121), Australia (103), and Germany (82) (opendatasoft, 2020).

Despite the enormous growth in the absolute number of open data platforms, recent years also brought about a shift in focus from the quantity to the quality of open data (Blank, 2019, 3). Since 2013, the OECD measures the availability, accessibility and re-use of open governmental data in its member and partner countries to assess the quality of their open governmental data policies (OECD, 2020). According to the OECD index, the majority of countries improved their open data policies in recent years. This development is reflected in the overall mean of the index scores across all countries: for 2017, the mean score is 0.54 (standard deviation 0.20) while the mean score for 2019 increased to 0.61 (standard deviation 0.14). For both years, South Korea is the uncontested leader in open governmental data policies. However, most EU member states are actually catching up quickly. France is the most advanced EU country both in terms of the quantity of available open governmental data (see above) as well as the quality of its open data policies.

Although the general trend has moved in the right direction and many improvements have been made with regards to governmental open data policies in recent years, governments still have to focus on continuous platform maintenance and user engagement.

2.3 From open data by governments towards data sharing by companies

Countries all over the world discover the huge potential of their data, which leads to the spread of open data initiatives started by governments and administrative organisations. While initially citizens were supposed to be the main beneficiaries of these data, companies are also encouraged to access the data provided by governments. Open data is also seen as a driver for innovation: companies can reuse data for their business models and entrepreneurs can use open data to create new business models (Lucke, 2012). The number of companies who utilise open data to generate profits is growing. For example, at least 31 percent of German companies whose processes involve data usage confirm using open data for the creation of products and services (Demary et al., 2019). South Korean firms are especially successful in using open data (OECD, 2020). More than 2,700 use cases are listed on the official South Korean data portal, illustrating how the open data has been used for national and international business ideas (Data.go, 2020).

While open governmental data is already widely published and used, sharing company data is still a niche phenomenon. Yet the enormous volume of data generated and stored by companies remains mainly unused (Heather Johnson, 2015). Operating on a smaller scale, companies are reluctant to publish their data and rather exchange it with partners or suppliers. Currently, 43 percent of data trading companies engage in bilateral data exchange, while only 6 percent trade data on open platforms (Demary et al., 2019).

For identifying reasons for reluctant data sharing behaviour of companies, we compare the characteristics of open governmental data and data shared by companies according to the five open data dimensions identified in chapter 2.1. Although a comparison between open governmental data and data shared by companies is not trivial, we expect that exploring its similarities will show which framework conditions for open governmental data can be applied for data sharing among companies, while exploring the differences will emphasise main hurdles for companies to share their data.

Strategic dimension
Some company data – like machine data or sensor data, which are non-personal and non-sensitive in terms of business secrets – have similar characteristics to open governmental data. This data could be made freely available for everybody without harming the supply side of the business. However, some company data contains personal or sensitive business information, which is not the case for open governmental data. Providing these types of data openly without any access limitation can harm the providing company or its customers, e.g. if data is accessed by competitors. Thus, full non-discrimination for all data types cannot be guaranteed in case of company data. In fact, a thorough examination of company data is necessary before it can be shared with external actors.
Upshot: In contrast to governmental data, not all company data can be provided open access. A strategic analysis is needed to identify data, which can be shared without harming a company's own business. This procedure can cause high information and transaction costs and requires an in-depth data expertise.

Economic dimension
On the one hand, similar to governmental data, data in companies are often a by-product, whose generation costs are consequently rather low. Thus, sharing this data for free or imposing a usage fee equal to its production costs without causing profit losses is possible. On the other hand, sharing data requires a governance structure in companies and charging an access fee can be a firm's business model. In fact, data as a product is a new business model, whose popularity increases steadily (Demary et al., 2019). In this case, usage costs are substantially higher than production costs. In addition, utility gains are located on the consumer side when governments share its data publicly, whereby companies can mutually profit financially or economically when sharing data with each other.
Upshot: While governments do not compete with each other, companies are profit-oriented organisations. Thus, the provision of data can be the core of a business model. In this case, data is not a public but a private good and cannot be shared with other actors, as it will harm the business model of the data owning company. In contrast, companies that generate data as a by-product can share their data and even mutually profit from the sharing process, e.g. through gaining and integrating information in the business process.

Legal dimension
Open licences commonly used for governmental data can also be applied to certain types of company data (mainly non-personal and anonymous data). The main difference between governments and companies sharing their data is the competition aspect. No competition exists between single governments when providing their data. In this case, data is a public good (non-

excludable and non-rivalrous). In contrast, companies can indeed be competitors when sharing data with each other. In this case, data has the properties of a club good (excludable and non-rivalrous). Other types of licences than open licences may be required for sharing this data. In this regard, only cartel law may restrict certain forms of data sharing among competitors.
Upshot: The legal framework is given for data sharing among companies when data is a public good. In cases when data reveal the characteristics of a club good or a private good, other licencing conditions are required for providing the data for which no established forms exist.

Conceptual dimension
Commonly owned or open standards can be applied both to governmental and company data. The Open Data Standards Directory developed by the McGill's GeoThink Research Group and the Johns Hopkins University's Center for Government Excellence contains more than 60 open standards for governments (Center for Government Excellence, 2020a). Meanwhile, some of these standards – like GTFS Realtime for real-time transit data – originate from companies (Google in this case), making them applicable for governments and companies (Center for Government Excellence, 2020b). There are some differences with regards to the enforcement of common standards. Compared to the governmental level, enforcement in the heterogeneous and decentralised company ecosystem becomes much more complex.
Upshot: The standards existing for governmental data can be applied to company data. Difficulties emerge in terms of establishing the state of the art for all companies willing to share their data.

Technical dimension
High quality of data is as important for companies as for governments. The IBM Big Data & Analytics Hub (2016) estimated the costs of poor data quality for US companies to be approximately $3.1 trillion a year. Thus, technical criteria in terms of data quality, machine-readability and continuity of access are relevant for both company data and governmental data.
Upshot: The technical criteria existing for governmental data are equal to company data which are provided to other actors.

Many similarities between governmental and company data illustrate that the fundamental conditions for sharing data among companies are already given when data has properties of a public good. Especially in terms of technical requirements, criteria used for providing open governmental data can work as a blueprint for sharing company data. As soon as data becomes a club or a private good for a company, hurdles for sharing data arise. The biggest disparities between the data types occur with regards to non-discrimination and usage costs. In addition, the legal concerns may cause insecurity for firms who are willing to share their data but do not know how. Thus, the different treatment of data as a good can explain why companies are still reluctant to share their data. Consequently, a deeper analysis of the data treatment and conditions for data sharing is needed to identify incentives and policy implications enabling companies to share their data to a higher extend. We will elaborate on this problem in the following chapter.

3 Incentives for data sharing among companies

In the previous chapter, we determined that governmental and company data have many similarities when data is treated as a public good for the owning company. When data is the core of a company's business model and profits are generated by e.g. selling data, sharing data becomes less attractive for companies. In this chapter, we aim to determine how data is treated in practice, focusing on Europe. We analyse to what extent and under which conditions data is currently shared among companies in the EU by examining the results of an open consultation by the EU commission. Subsequently, we use the data sample to identify the incentives European companies seek for to increase the amount and intensity of data sharing.

3.1 Data sample and research method

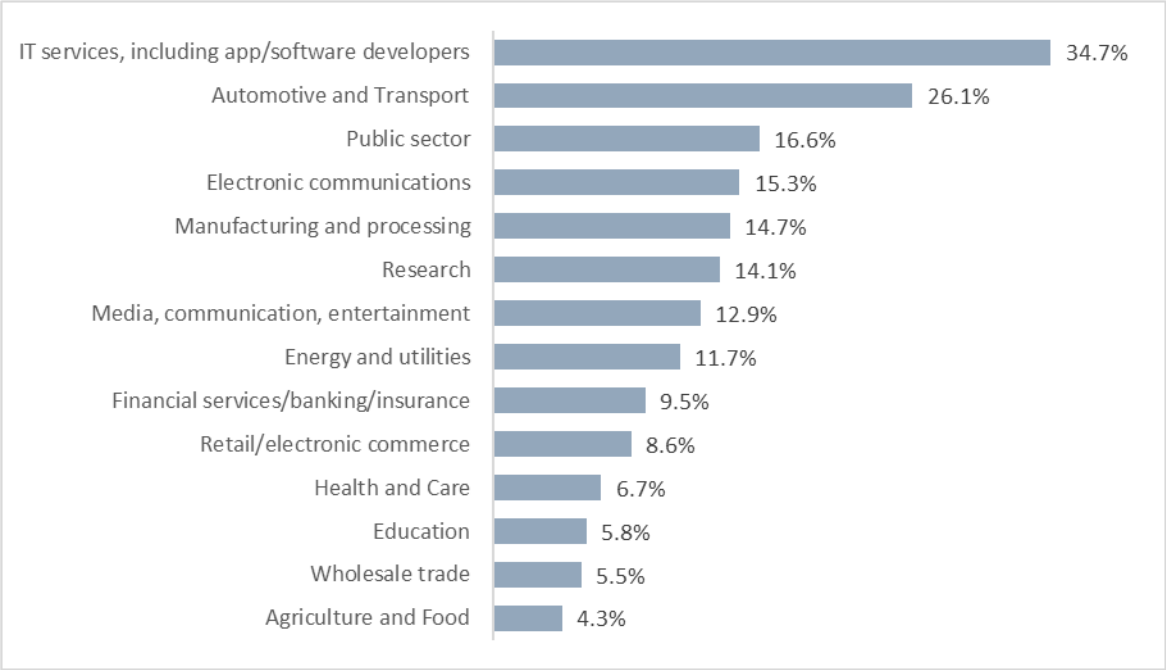
The following analysis builds upon a public consultation by the EU Commission. Through online public consultations EU citizens can express their views on the scope, priorities and added value of EU action for new initiatives, or evaluations of existing policies and laws (European Commission, 2020b). The chosen online consultation on Building the European Data Economy was conducted from 10 January 2017 to 26 April 2017 (European Commission, 2017). One of the main objectives of the consultation was to collect information on the extent of data sharing and trading as well as the existing barriers for the sharing/trading process. Analysing the responses of the participants, we seek to identify which incentives companies in the European Union need in order to share their data with external stakeholders.

The original dataset of the EU commission reveals 380 responses. The questionnaire consisted of multiple-choice, single-choice as well open ended questions. In a first step, duplicates (in columns second name and/or organisation) were removed, such that a total of 377 responses remained in our final sample. Relevant questions in the data sample were identified and analysed. Furthermore, we assume that participating companies at least show interest for data-related questions or already engage in data-driven businesses. Our analysis serves as an illustration of the status quo of a certain fraction of European firms. However, we assume that the results of this analysis may nevertheless help identifying possible barriers to data exchange by business actors and might encourage companies to share their data with each other.

Companies from all sectors were encouraged to participate on the consultation. As Figure 3-1 shows, all sectors are present in the sample. Most dominant are IT services and software developers (34.7 percent). This is not surprising when considering the fact that companies in this sector primarily work with data and information of any kind. The second highest fraction of participants belong to the automotive and transport sector (26.1 percent).

Figure 3-1: Sectors of participants

Multiple selection possible; n=326

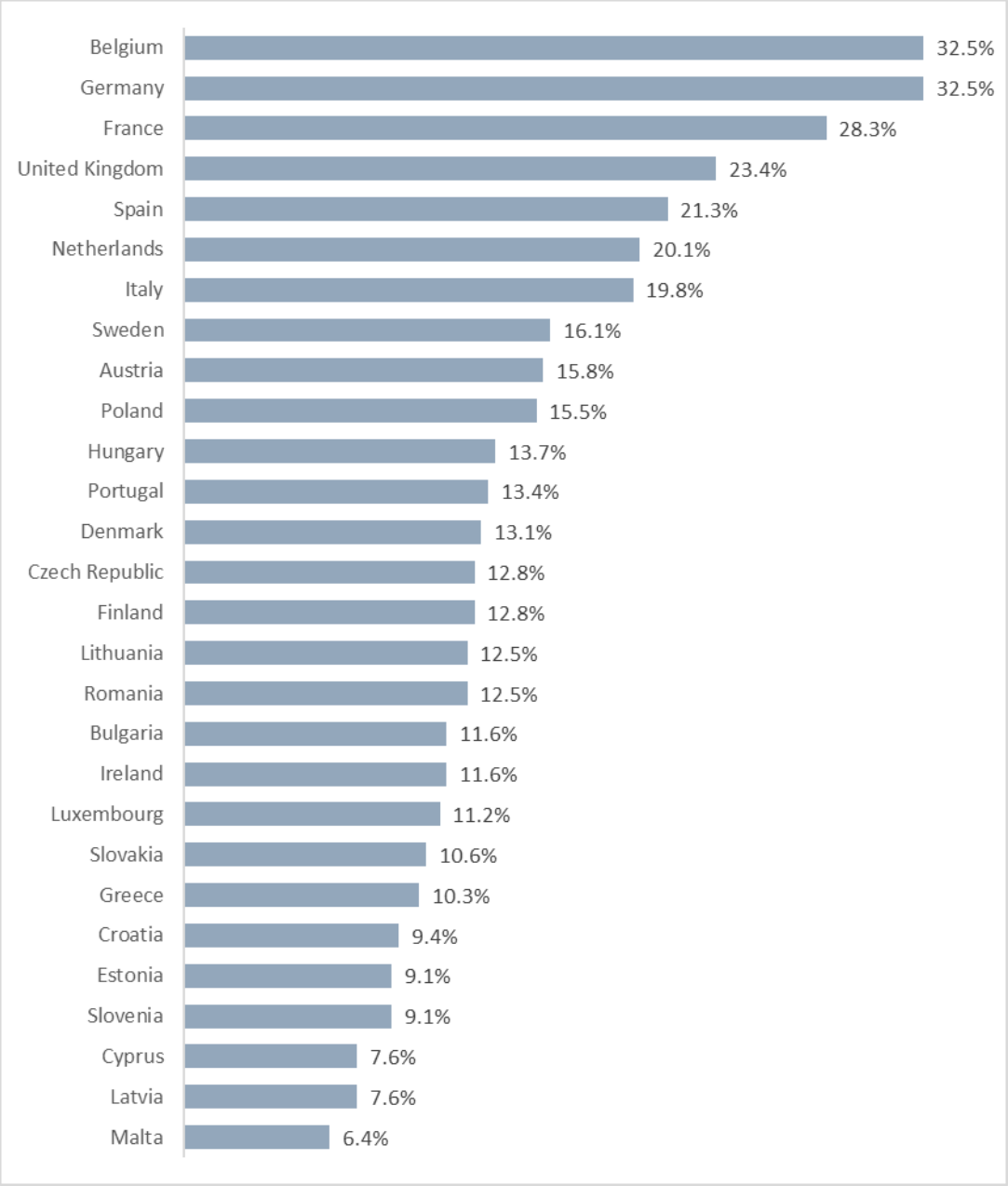


Source: EU Commission, 2017; own representation

Figure 3-2 shows that the final sample contains participants from all 27 EU-countries (plus the United Kingdom (UK), which was still an EU member state in 2017). Companies operating in the UK were left in the final sample as they were treated more or less as an EU member state during the transition period until December 31, 2020 (European Union, 2019b, 2019a). Most participating companies and organisations operate their business in Belgium and Germany (both 32.5 percent) followed by France (28.3 percent), the UK (23.4 percent), and Spain (21.3 percent). The high share of companies from Germany, France, Spain or the UK is not surprising when taking into account that they belong to the largest countries in the EU in terms of population size and surface area. The sample bias towards Belgian companies can be explained by the fact that the headquarters of many EU institutions – including the EU Commission – are located in Brussels. To sum up, the survey provides a useful diversity in terms of sectors and countries the participating companies or organisation operate in.

Figure 3-2: Place of business

Place(s) of operation of business/organisation; multiple selection possible; n=329



Source: EU Commission, 2017; own representation

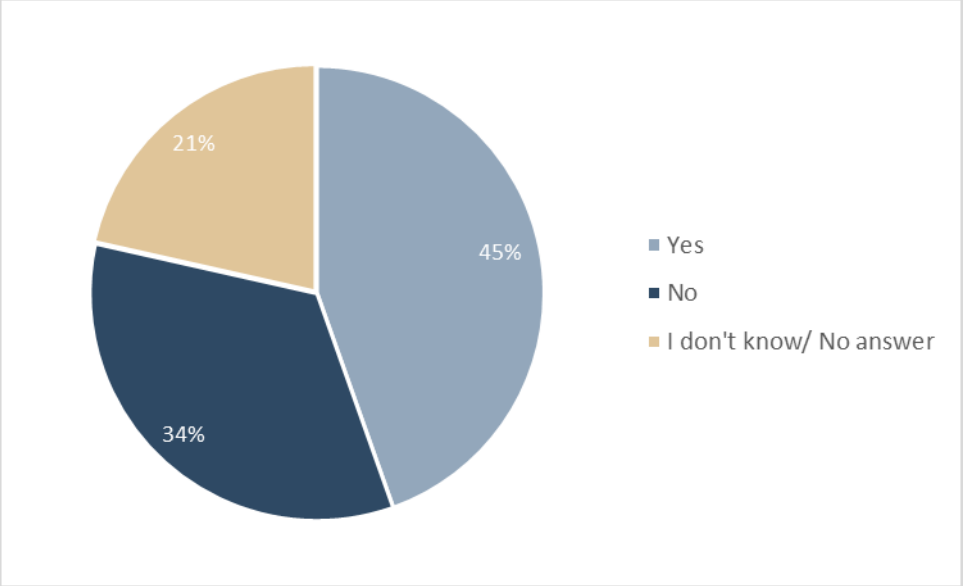
3.2 Status quo in data sharing

In order to analyse the incentives required to intensify data sharing and data trading processes between companies, we first take a glance at the status quo for participants. By analysing the answers in the online consultation, we seek to identify the firms that are interested in data sharing. First, we evaluate how many companies rely on purchased data. Figure 3-3 shows that 45 percent of the participants depend on data resources acquired from other actors, while only 34 percent are independent from external data. The fact that every third company does not acquire external data for its business does not necessarily imply that these companies do not

use data. In fact, they could be a provider of data or build their business model on company's internal data.

Figure 3-3: Data dependence

Do you currently depend to a significant extent on data resources that you acquire from others (for products or services you offer, for your internal business processes)?; n=377

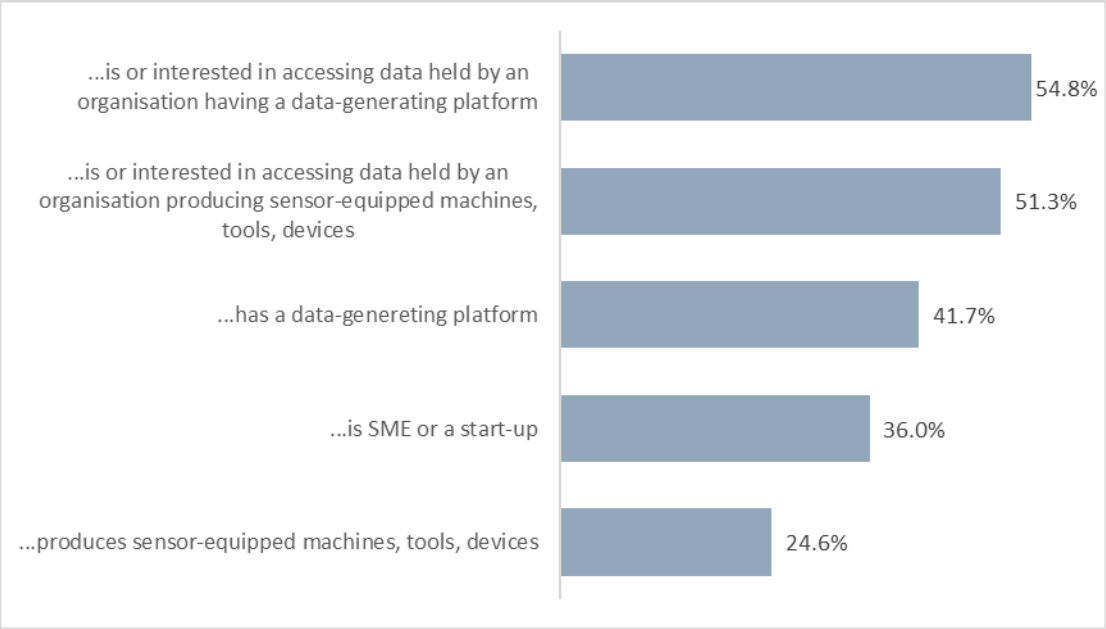


Source: EU Commission, 2017; own representation

The majority of the participating firms is already accessing data held by an organisation which makes business using internet-based platforms or is interested in doing so (54.8 percent; see Figure 3-4). Almost as many firms are accessing data held by an organisation which has significant business in the production and market commercialisation of sensor-equipped machines, tools, devices or are interested in doing so (51.3 percent). Thus, the fraction of the firms accessing or wishing to access data from third parties in the data sample is higher than the fraction of firms proving data generated on a platform (41.7 percent) or by sensors during the production processes (24.6 percent). Remarkably, 36 percent of the participants indicate themselves as small/medium enterprises or start-ups, which reveals that the interest in data and data-driven services is not just subject to large enterprises but also affects small businesses or start-ups.

Figure 3-4: Data context: my organisation…

Which (if any) of these statements apply to you?; multiple selection possible; n=228

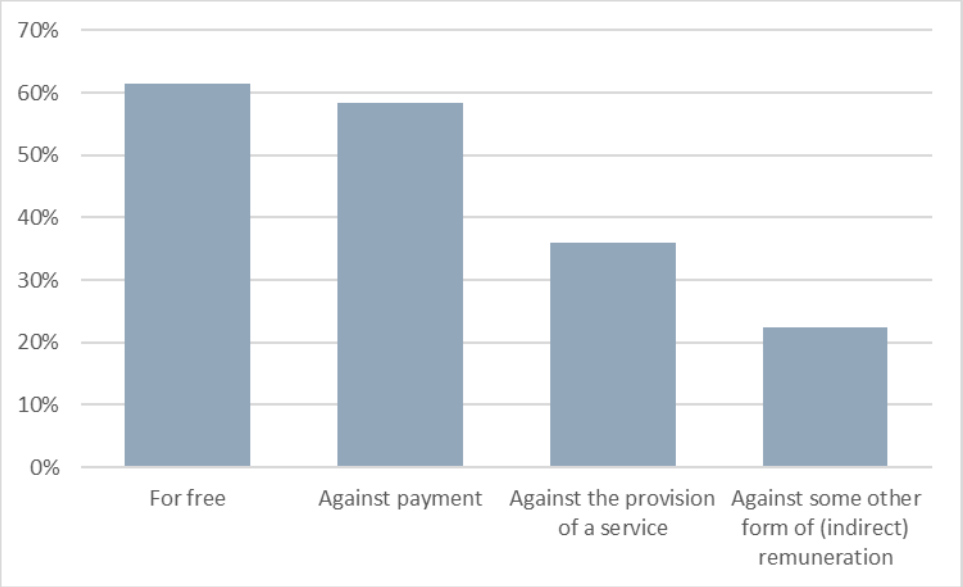


Source: EU Commission, 2017; own representation

Interestingly, when acquiring data from third parties, most companies access such data free of charge (61.5 percent), while 58.3 percent of the companies purchase external data (see Figure 3-5). This fact implies that frequently data is not subject to a business model for companies but is shared freely like governmental data. More than one third of respondents receive data by providing some sort of service (35.9 percent): data analysis as a data-driven business model has been gaining popularity in recent years. For example, producers of machines equipped with sensors may offer additional services, such as identifying malfunctions or predicting technical replacements or maintenance services. This service is called predictive maintenance or smart production.

Figure 3-5: Acquisition conditions for data

If you are acquiring non-personal or anonymised data produced by others, what are the remuneration conditions for accessing the data?; Multiple selection possible; n=156

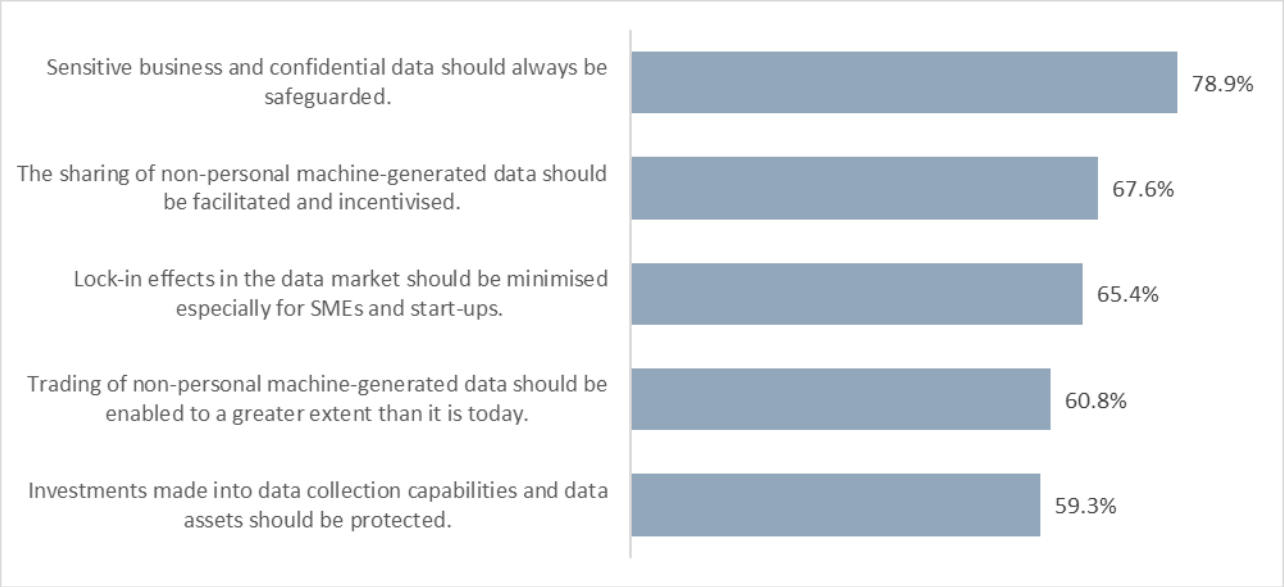


Source: EU Commission, 2017; own representation

Although a large proportion of the companies that participated in the consultation already engage in data sharing to a greater extent, they confirm that data sharing processes can still be improved. Figure 3-6 shows that almost 80 percent of the respondents state that sensitive and confidential data should always be safeguarded. As we discussed previously, this kind of data has to be excluded from data sharing between companies. In contrast, the sharing of non-personal and non-sensitive machine data should be facilitated and incentivised. This idea was supported by 68 percent of the consultation participants and almost 61 percent presume that, in addition, trading such data has not reached its full potential yet and should be extended. More than 65 percent of the respondents agree that the portability between different data markets is an important feature for data sharing. Especially small and medium-sized companies and start-ups would otherwise suffer if the switching costs were too high due to lock-in effects on data platforms. The portability between data market platforms ensures competition conditions on the market and prevents the development of monopolistic market structures. Finally, 59 percent of the participants emphasise the importance to protect investments made in data collection and data assets in general.

Figure 3-6: Data Sharing and Data Trading

To what extent do you agree with the following statements? Participants who answered "to some extent" or "to a great extent"; n=285



Source: EU Commission, 2017; own representation

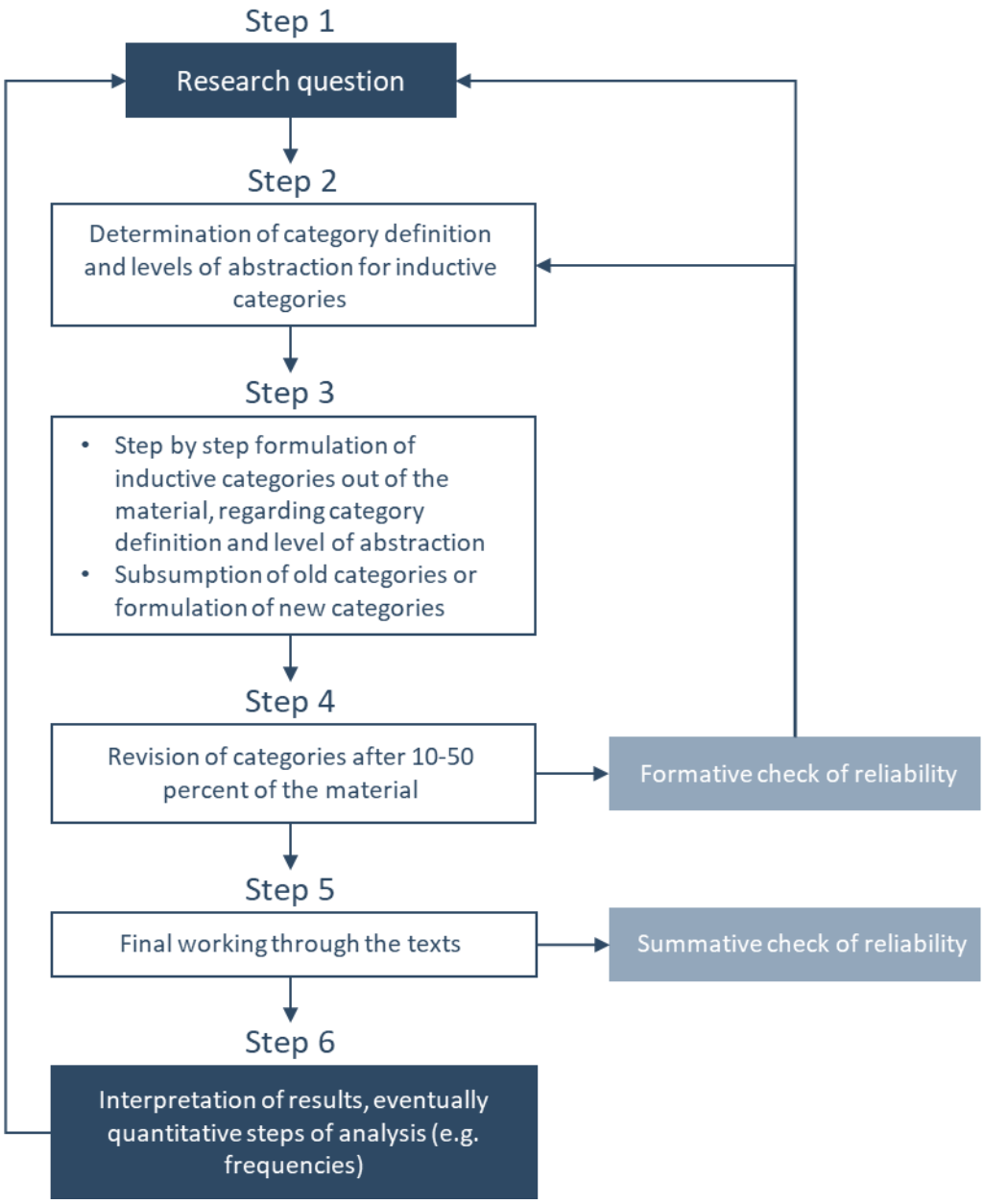
The results confirm the presumption that most companies participating in the consultation have a strong connection to data-driven businesses. In particular, the interest in increased sharing of non-personal and non-sensitive data is very high. Moreover, the fact that more companies acquire data for free or for other kinds of renumeration than financial payment, indicates a high fraction of firms in the sample treating data as a public good and thus, are less reluctant towards data sharing. Consequently, we can safely assume that the readiness to share data already exists for most participating companies and given the right incentives, would be willing to share more of their data in the future.

3.3 Incentive analysis

In the previous chapter, we described the status quo in data sharing behaviour of EU companies. We identified that some companies already share data, while some build their business models on data provided by third parties. However, there is still a vast amount of unused potential of data sharing. Hence, the aim of this chapter is to identify sufficient conditions needed by European companies to foster the readiness to share company data with other actors.

We apply a qualitative content analysis developed by Mayring (2015), whose general structure is explained in the following. In our analysis, we use the inductive category development method, where the definition of the inductive categories takes place iteratively in the process of the analysis. As Mayring (2015) indicates, the advantages of this method are the closeness to the original subject (responses of participants in this case) and its unbiased results. The procedure consists of six steps (see Figure 3-7).

Figure 3-7: Steps of the inductive category development



Source: own representation based on Mayring, 2015

In the first step, the research question is formulated. For our analysis, the determined research question is: “Which incentives are needed by European companies to foster their readiness to share company data with other actors?”.

Then, the selection criteria and the level of abstraction follow, derived from the theoretical background and the research question. In this step, one defines the subject of the text analysis and how detailed the analysis will be. The selection criteria determine categories of the analysed text. Accordingly, the subject of our analysis corresponds to all responses to the open ended question “What kind of incentives would make you share data with a wider range of economic

operators?" from the examined consultation. The responses were free texts written by participants. In total, 84 participants gave detailed answers to this question. Identical responses were removed to eliminate the bias created by interest groups of participants who purposely gave the same answer to the question. Subsequently, the final sample consisted of 81 responses. In order to harmonise the given sample, we translated the responses in languages other than English via the machine translation tool DeepL or alternatively a combination of Google Translate and Systran for languages not available in DeepL. We decided to apply a manual coding procedure on the data sample. The coding implied summarization of each response in key words, which built the temporary categories.

Based on the defined criteria, the text is processed by assigning categories to certain paragraphs or sentences in step 3. Categories can be short summaries or key words of a single text passage, for example. During this process, categories remain temporary and can be re-formulated or extended step by step. The aim of this procedure is to create a clear structured overview of the analysed text by coding, such that the content is represented in built categories. The coding process should first be applied to an ex-ante determined fraction of the whole text (10 to 50 percent). In our analysis, the first codification divided the sample into positive, neutral and negative statements based on the sentiment of each response. As a result, we classified 62 responses as positive, 9 responses as neutral and 10 as negative. In the next step, the responses were coded based on its content, resulting in temporary categories. For the main analysis, we first focused only on positive responses.

Afterwards, a revision for categories in step 4 begins. In this step, the categories are revised and summarised to broader categories, if necessary, while the reliability of the remaining categories has to be ensured. Reliability is given when the existing categories still represent the main statements of the text. If the reliability is not sufficient the procedure returns to the first or second step. We applied the first revision phase, in which we merged some categories or removed them entirely, after coding half of the responses. The double-check of the categories by corresponding authors assured the reliability of the remaining categories.

In step 5, the remaining text is processed, categories are summarised to main categories (if necessary) and undergo the last reliability check according to the final categories. In our analysis, the remaining categories were grouped thematically into six categories, which represent six different incentives for data sharing: identification of data sharing potential, realisation of business opportunities, introduction of an appropriate legal and contractual framework, implementation of reciprocal data sharing, realisation of financial benefits and serving customers' needs. The intercoder reliability (double-check of the criteria by the co-author) ensured the quality criteria of the analysis to be fulfilled.

Table 3-1: Identified incentives for data sharing

Incentive	Explanation
Identification of data sharing potential	Participants express difficulties in finding the value of data sharing. They need precise explanations why their company should provide access to their data or why they should obtain data from other companies.
Realisation of business opportunities	Data sharing can be incentivised through added value created due to access to different data. Business opportunities resulting from data sharing are manifold. Better performance, innovative processes within own company, development of new markets and acquisition of new customer segments are some examples named by participants.
Introduction of an appropriate legal and contractual framework	Participants demand clear laws and rules for the whole data sharing process starting with data generation and ending with data delivery. The legal conditions should be transparent, practical, and non-discriminatory. Participants frequently mention licencing rules as essential requirement for incentivised data sharing.
Implementation of reciprocal data sharing	Some companies are ready to share their data with other stakeholders if the sharing process is reciprocal, i.e. other stakeholder also provide their data in return. The exchange of data can be within partners, intrasectoral or intersectoral.
Realisation of financial benefits	Companies would share their data if the financial benefits substantially outweighed the sharing costs. Financial benefits include payment for provided data or reduction of costs for access to data sources.
Serving customers' needs	Some participants claim that company data should only be shared if customers express a need or an explicit permission to do so. In this case, mainly customer data or data related to customers are relevant.
Other	Among other incentives, reliable cyber security, further scientific research, establishing ecological standards and sharing of information instead of raw data were named.

Source: own representation

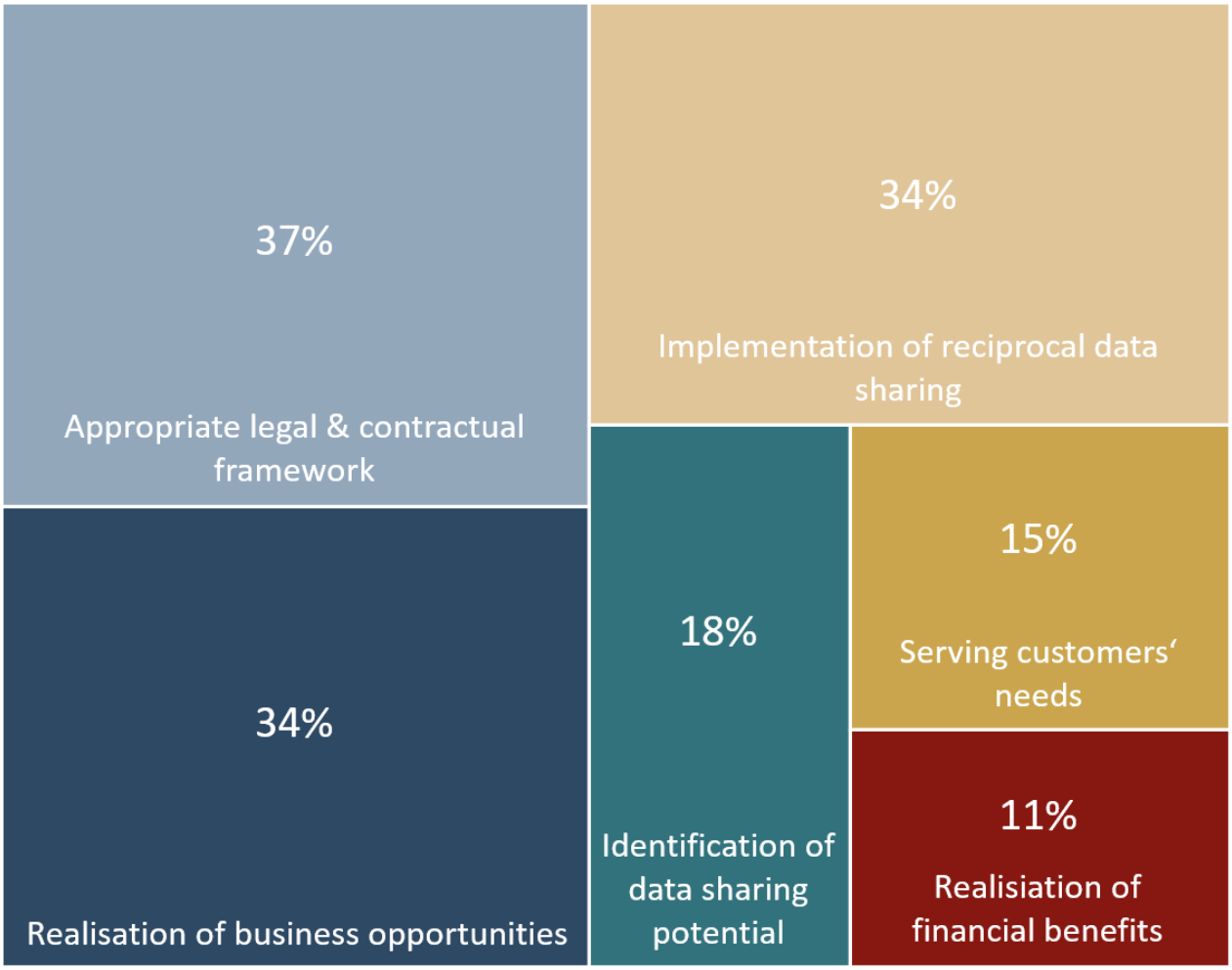
In the conclusive step of the qualitative text analysis, the interpretation of the results follows based on the given research question. The interpretation may contain a quantitative analysis of the results (e.g. descriptive statistics on the categories). Table 3-1 summarises the identified incentives for data sharing expressed by the participants of the survey.

We identified that most participants (37 percent, see Figure 3-8) indicate the need for an appropriate legal and contractual framework, which indicates legal uncertainty as the most urgent problem to be solved. They state that the current regulatory environment impedes stakeholders' understanding of what kind of data can be shared and under which conditions. In particular, clear sanctions and remedies against inappropriate and unethical use of data should ensure data usage only for purposes defined in a licence. For instance, one participant demands a "clear ban on potentially discriminatory behaviour from some economic actors" and another participant considers "practical guidelines on data anonymization process" helpful. Further possible instruments for fair use of data are an obligation for data transfer as well as traceability or common standards. Another indicated uncertainty is connected to ownership rights: more data sharing can be achieved by the "creation of a non-exclusive, flexible and extensible ownership right in data". Some participants complain that not all stakeholders have equal legal rights in terms of data exchange: there must be a "level playing field" between companies providing public services and private companies. The regulation should be located at the EU level to prevent a "race to the bottom between member states". Besides, the created legal framework should be as transparent as possible. All stakeholders should possess the same clear information on their legal rights.

More than one third of the respondents sees the reciprocity of data sharing as an important incentive (34 percent). Thus, all ecosystem members should provide access to their non-sensitive data to ensure mutual benefits and prevent some actors from free-riding. For this reason, a participant states: "it is fairer if everyone shares their data equally". Other participants confirm to be incentivised to share their data when prospective collaboration is given or if synergies cannot be realised without data sharing. Other respondents require the "same vision for data usage" and a cross-sectoral discussion to foster data sharing. For achieving reciprocity, the following applies: reciprocity can be more easily maintained in bilateral or small, closed groups of companies. In ecosystems with a larger number of members, however, the heterogeneity of the firms and sectors as well as their international orientation constitute typical obstacles for enabling reciprocal data sharing.

Figure 3-8: Frequency of identified incentives

Fraction of responses belonging to each category, n=62



Source: own representation

Another important incentive mentioned by 34 percent of the participants is the realisation of business opportunities with shared data. Hence, they assume to share their data more frequently if they benefit economically and commercially. The range of listed possibilities of how to generate business value varied considerably. Some participants assume that data accessed from other actors could improve their own performance, maintain growth, and facilitate the development of new data-based services and products. Also, analytics-as-a-service business models are mentioned by some participants. By providing their own data to other companies specialised in analytics, new insights in data and decision support can be achieved. Other participants see the potential to share benefits between business partners by combining data from different sources, for example through "joint product development", "cross-selling" or "realignment of cross-company business processes". Besides the disruptive potential of data such as entry of new markets, innovation or the acquisition of new customers were mentioned as decisive incentives. For other companies, the "ability to fully participate in a data economy" creates added value through transparency and the availability of diverse data types. In spite of obvious

benefits of data sharing, some respondents highlighted that economic benefits are only possible if the created value exceeds the costs of sharing, while risks of losing trust and reputation should be also taken into account.

Though there are obviously many existing possibilities of how data can enhance business performance, 18 percent of the participants still claim the benefits of data sharing should be clarified and promoted for all members of a data ecosystem. Some companies ask how they can generate profits by using their own or external data or state to have no demand for data.

Serving customers' needs (15 percent) and creating financial benefits (11 percent) were mentioned less frequently compared to the other incentives. For some companies, the customer is the final decision maker. Thus, no data can be shared without her explicit permission. In the first place, data sharing should serve customer interests and customer benefits. Other participants confirm that they would start sharing data if their customers would demand it. If privacy and security requirements of customers cannot be satisfied, though, no incentive to share data is given. Meanwhile, just a few respondents highlight financial benefits from data sharing as an important incentive. However, for some of them it was the only required incentive. These companies supply their data solely for a reasonable price. Providing open access to the data is obviously not the preferred option for these firms, they rather treat data as a product. Another participant demands lower access costs for data in order to incentivise the sharing processes.

Last, we also examined neutral and negative responses to the question. Some participants declare the decision to not share their data to be a business rule. Other companies do not share data as they do not see any benefits in doing so, neither for their own company nor for external actors. One participant states that "generating data is resource-intensive" and data quality is too high "to provide it free of charge". Another respondent criticises the current development at the EU level to go into the wrong direction. Some companies hesitate to share data because more clarity on this subject is needed from the EU Commission or because they first want to monitor possible business cases which are suitable for data exchange, which supports the incentives derived previously. One participant states that no fair play between large companies and start-ups exist in her country, which impedes data sharing. Only a few participants reply that no incentives are required to increase data sharing among companies and that the created conditions have already caused a fast increase in data sharing.

To sum up, the number of participants answering the analysed question and the number of suggestions indicate a high interest to share data among participating European companies. However, many respondents also clarify that there is still a need for improving the incentives structure that encourage more companies to facilitate to access company data.

4 Recommendations for action

In the previous chapter we described the status quo of data sharing and identified incentives to facilitate data sharing processes between companies. Taking the results into account, we derive five distinct policy recommendations in this chapter. We are convinced that this endeavour can only be successful if all actors are willing and able to work together on developing an effective data ecosystems.

Create common contract rules and standards

Enhancing data sharing among companies requires clear and applicable legal rules for data. Taking the public or club good character of data - characterized by non-rivalry in consumption – into account, a closer look reveals that an explicit property right for data does not exist. In this regard, the economic literature shows that introducing a new property or intellectual property right for data is not reasonable (Rusche/Scheufen, 2018; Duch-Brown et al., 2017; Kerber, 2016).

While the usage of personal data is comprehensively regulated by the General Data Protection Regulation (GDPR), there is hardly any regulation regarding non-personal data apart from general rules of contract or cartel law. Nevertheless, the current legal framework already enables selling or sharing of data by means of individual contractual agreements between the parties involved (Fries/Scheufen, 2019). The simple fact of holding data enables the holder of data to contractually grant certain property rights. As such, the data holder can restrict access to data by specifying a certain time frame by giving access to specific parts of the data set or even by specifying explicit responsibilities and hence legal liability for securing the data. Regarding the latter aspect, the German criminal law emphasizes the securing of data as a prerequisite for liability and claims for damages. Interestingly in this context, Shen et al. (2020) suggest the use of blockchain technologies to solve security and privacy issues.

Despite existing legal rules for enabling selling and sharing of data, prohibitive information and transaction costs of individually negotiating contracts may likely cause levels of data sharing that are lower than the socially optimal level. Consequently, policy instruments explicitly directed towards lowering the information and transaction costs of data sharing are required from an economic point of view. A provision of contractual principles and standards could help in fostering data sharing between firms (Fries/Scheufen, 2019), e.g. by offering default rules for data sharing contracts or by developing a new licensing scheme explicitly directed towards the sharing of data between firms. In this regard, the EU data strategy may only be a first step in this direction (European Commission, 2020a).

Facilitate cooperation between different actors

Currently, data is mostly shared in bilateral contract agreements (Demary et al., 2019). Long-term collaborations or even data networks with several partners are rare. To improve the exchange of data between companies, cooperation between traditional companies and start-ups should be empowered (Engels/Röhl, 2019). Combining the innovation potential of start-ups and the financial power and infrastructure of traditional companies enables new data-driven business opportunities. Also, the importance of the cooperation between business and science

should not be neglected. Scientific expertise provides companies with the required technological, economic, and sociological input, while business actors show significant developments and current important issues for research work in turn.

Companies need help finding suitable partners for the generation of business ideas. The network of Digital Innovation Hubs (DIH) initiated by the EU is an excellent example of how SME, corporates, entrepreneurs and scientists can connect in order to create innovation impulses in different areas and a rise of new business models or even markets, which in turn can manifest the EU as a strong digital competitor in the world. Currently, 635 of such DIHs in different European countries, including non-EU-members, have already registered in the community (European Commission, 2020c). The initiative is still in process and it remains to be seen whether it will succeed. The development of intrasectoral data spaces as it has been announced in the EU data strategy will build the next step in this direction.

Provide a trustworthy technological infrastructure

Data ecosystems are not very common yet but are emerging as the basis for business models of the future. For reciprocal data sharing between a range of business actors, a proper technology is needed to ensure trust and fair conditions for all participants. The International Data Spaces (IDS) is a reference architecture for sovereign data exchange between companies. It guarantees "data sovereignty by an open, vendor-independent architecture for a peer-to-peer network which provides usage control of data from all domains" (IDS, 2020). Currently, more than 100 companies from different industries and institutions are members of the initiative (ibid.). The IDS reference infrastructure is still developing. Hence, further support and research are needed in this area to establish the IDS as the international market standard for the sovereign use of data and encourage more companies to participate in this initiative.

Another example for a promising project is GAIA-X – also known as "European cloud", a federated, open data infrastructure for Europe (BMWi, 2020). The main goal of this initiative is to create a transparent and sovereign data and infrastructure ecosystem for business innovation, based on open standards and the European legislative. While seven EU-countries already participate in developing GAIA-X, other European countries are expected to follow. These initiatives are both steps in the right direction. It remains to be seen whether they will be successful and achieve the anticipated goals.

Develop a data strategy

A lot of companies are still not or barely digitalised such that they rely on analogous business structures. For now, they have little incentives to change their business models as long as they remain profitable. However, these companies jeopardise their competitive position in the future by giving other companies the edge in digital innovation. Missing the right moment to digitally transform one's own business could mean significant financial losses and even bankruptcy in the worst case, as the example of the once leading technology company Kodak shows.

One of the reasons why many companies hesitate to digitalise their processes is that data-driven innovation requires high up-front investments and a risk-taking attitude. The impulses to inno-

vate should come from the high level accompanied by an appropriate data strategy. The development of a data strategy is a decisive first step towards data-driven innovation. When companies define a data strategy, they engage with the data available to them and the potential benefits that can be generated. A data strategy should include a long-term business plan and consider current and assumed future developments of the market. Thereby, a data strategy does not need to concentrate on a single business case but may include a range of business ideas that can potentially be implemented. These business ideas should all focus on the customer needs. Companies should identify the needs and pains of their customers to develop data-driven services and products providing solutions.

Determine the value of data

In order to share data or to build a new business model around it, companies should know the value of their data. Knowing the value of their own data helps companies to quantify benefits and costs connected to their data. Especially if data is to be traded (for example on a data marketplace), a market price is required. As the trend goes towards data being treated as a company resource, methods are needed to determine its value. As for now, only few companies and researchers deal with this issue, though it is essential for further development of data sharing initiatives. The most common approaches are cost-based, utility-based and market-based valuation (Krotova et al., 2019). However, no universal methods and tools for determination of data value currently exist. Thus, clear rules and standards are needed in this matter, accompanied by further research.

References

Blank, Marit, 2019, Open data maturity report 2019, Open data maturity report 2019, [Luxembourg]

Center for Government Excellence, 2020a, Open data standards directory, https://datastandards.directory/ [2020-07-16]

Center for Government Excellence, 2020b, Open data standards directory API. Real-time transit, https://datastandards.directory/real-time_transit [2020-07-16]

Data.go, 2020, Open Data Portal, https://www.data.go.kr/tcs/puc/selectPublicUseCaseListView.do [2020-07-16]

Delhaes, Daniel, 2020, Deutsche Autoindustrie erwägt, ihre Datenschätze zu bündeln, in: Handelsblatt, Sep 7

Demary, Vera et al., 2019, Readiness Data Economy. Bereitschaft der deutschen Unternehmen für die Teilhabe an der Datenwirtschaft, Köln

Duch-Brown, Nestor / Martens, Bertin / Mueller-Langer, Frank, 2017, The economics of ownership, access and trade in digital data, JRC Technical Reports, JRC Digital Economy Working Paper 2017-01, Seville, Spain

Engels, Barbara / Röhl, Klaus-Heiner, 2019, Start-ups und Mittelstand. Potenziale und Herausforderungen von Kooperationen = Start-ups and Germany's "Mittelstand" (SMEs) : the potential and the challenges of collaboration, Köln

European Commission, 2017, Public consultation on Building the European Data Economy, https://ec.europa.eu/digital-single-market/en/news/public-consultation-building-european-data-economy [2020-07-16]

European Commission, 2020a, A European strategy for data, Brussels

European Commission, 2020b, Consultations, https://ec.europa.eu/info/consultations_en?order_by_status=All&field_core_topics_target_id_entityreference_filter=369 [2020-07-16]

European Commission, 2020c, Smart specialisation platform. Digital Innovation Hubs, https://s3platform.jrc.ec.europa.eu/digital-innovation-hubs-tool [2020-07-22]

European Union, 2019a, Agreement on the withdrawal of the United Kingdom of Great Britain and Northern Ireland from the European Union and the European Atomic Energy Community. 2019/C 384 I/01

European Union, 2019b, Political declaration setting out the framework for the future relationship between the European Union and the United Kingdom. 2019/C 384 I/02

Federal Ministry for Economic Affairs and Energy (BMWi), 2020, GAIA-X: Driver of digital innovation in Europe. Featuring the next generation of data infrastructure, https://www.bmwi.de/Redaktion/EN/Publikationen/gaia-x-driver-of-digital-innovation-in-europe.pdf?__blob=publicationFile&v=8 [2020-07-21]

Free Software Foundation Europe, 2020, What is Free Software, https://fsfe.org/freesoftware/freesoftware.en.html [2020-07-16]

Fries, Martin / Scheufen, Marc, 2019, Märkte für Maschinendaten: Eine rechtliche und rechtsökonomische Standortbestimmung, in: MMR, Vol. 11, pp. 721–726

Fritsch, Manuel / Krotova, Alevtina, 2020, IW-Report09/20 - Wie datengetrieben sind Geschäftsmodelle in Deutschland? Analyse des Status quo

Google Trends, 2020, Google Trends. Open Data, https://trends.google.de/trends/explore?date=all&q=%2Fm%2F0119t0j6 [2020-07-22]

Heather Johnson, 2015, Digging up dark data: What puts IBM at the forefront of insight economy, https://siliconangle.com/2015/10/30/ibm-is-at-the-forefront-of-insight-economy-ibminsight/ [2020-07-16]

Huyer, Esther / van Knippenberg, Laura, 2020, The Economic Impact of Open Data. Opportunities for value creation in Europe, Brussels

IBM Big Data & Analytics Hub, 2016, Extracting business value from the 4 V's of big data, https://www.ibmbigdatahub.com/sites/default/files/infographic_file/4Vs_Infographic_final.pdf

International Data Spaces Association (IDS), 2020, About the International Data Spaces Association, https://www.internationaldataspaces.org/our-approach/#about-us [2020-07-21]

Janssen, Marijn / Charalabidis, Yannis / Zuiderwijk, Anneke, 2012, Benefits, Adoption Barriers and Myths of Open Data and Open Government, in: Information Systems Management, Vol. 29, No. 4, pp. 258–268

Jetzek, Thorhildur, 2016, Managing complexity across multiple dimensions of liquid open data: The case of the Danish Basic Data Program, in: Government Information Quarterly, Vol. 33, No. 1, pp. 89–104

Kalampokis, Evangelos / Tambouris, Efthimios / Tarabanis, Konstantinos A. (eds.), 2013, Linked Open Government Data Analytics

Kerber, Wolfgang, 2016, A New (Intellectual) Property Right for Non-Personal Data? An Economic Analysis, in: Gewerblicher Rechtsschutz und Urheberrecht, Internationaler Teil (GRUR Int), No. 11, pp. 989–999

Krotova, Alevtina / Rusche, Christian / Spiekermann, Markus, 2019, Die ökonomische Bewertung von Daten. Verfahren, Beispiele und Anwendungen = The economic evaluation of data : procedures, examples and applications, Köln

Kuzev, Pencho / Dapp, Marcus M. / Balta, Dian / Palmetshofer, Walter / Krcmar, Helmut (eds.), 2016, Open Data. The Benefits. Das volkswirtschaftliche Potential für Deutschland, Sankt Augustin, Berlin

Lämmerhirt, Danny / Rubinstein, Mor / Montiel, Oscar, 2017, The State of Open Government Data in 2017. Creating meaningful open data through multi-stakeholder dialogue

Lucke, Jörn von (ed.), 2012, Entdeckung, Erkundung und Entwicklung 2.0. Open government, open government data und open budget 2.0, Berlin

Mayring, Philipp, 2015, Qualitative Inhaltsanalyse. Grundlagen und Techniken, Weinheim, Basel

OECD, 2018, OECD Open Government Data Report, Paris

OECD, 2019, Enhancing Access to and Sharing of Data:. Reconciling Risks and Benefits for Data Re-use across Societies, Paris

OECD, 2020, OECD Open, Useful and Re-usable data (OURdata) Index: 2019, http://www.oecd.org/gov/digital-government/ourdata-index-policy-paper-2020.pdf [2020-07-21]

Open Knowledge Foundation (OKF), 2015, The Open Definition, http://opendefinition.org/ [2020-07-16]

Open Knowledge Foundation (OKF), 2020, Mission. Our mission - an open world, where all non-personal information is open, free for everyone to use, build on and share; and creators and innovators are fairly recognised and rewarded., https://okfn.org/about/ [2020-07-16]

opendatasoft, 2020, Open Data Inception, opendatainception.io [2020-07-10]

Ostrom, Elinor, 1990, Governing the commons. The evolution of institutions for collective action, Cambridge

Ostrom, Elinor, 2009, A general framework for analyzing sustainability of social-ecological systems, in: Science (New York, N.Y.), Vol. 325, No. 5939, pp. 419–422

Richter, Heiko / Slowinski, Peter R., 2019, The Data Sharing Economy: On the Emergence of New Intermediaries, in: IIC - International Review of Intellectual Property and Competition Law, Vol. 50, No. 1, pp. 4–29

Rusche, Christian / Scheufen, Marc, 2018, On (intellectual) property and other legal frameworks in the digital economy: An economic analysis of the law, in: IW-Reports, No. 48

Shen, Meng / Zhu Liehuang / Xu Ke, 2020, Blockchain. Empowering secure data sharing, [S.l.]

Sunlight Foundation, 2017, Ten Principles For Opening Up Government Information, https://sunlightfoundation.com/policy/documents/ten-open-data-principles/#:~:text=The%20principles%20are%20completeness%2C%20primacy,licensing%2C%20permanence%20and%20usage%20costs. [2020-07-16]

Tammisto, Yulia / Lindman, Juho (eds.), 2012, Definition of Open Data Services in Software Business

van den Eynden, Veerle / Bishop, Libby, 2014, Sowing the seed: Incentives and motivations for sharing research data, a researcher's perspective, https://www.knowledge-exchange.info/event/sowing-the-seed [2020-07-21]

Wewer, Göttrik, 2020, Offene Daten (Open Data), in: Klenk, Tanja / Nullmeier, Frank / Wewer, Göttrik (eds.), Handbuch Digitalisierung in Staat und Verwaltung, Wiesbaden

Yu, Harlan / Robinson, David G., 2012, The New Ambiguity of 'Open Government', in: UCLA Law Review Discourse, Vol. 59, pp. 178–208