



# Progress Report

Expert Group for the Observatory on the Online Platform Economy

# Work stream on Data



# Contents

|                                                                                                                      |           |
|----------------------------------------------------------------------------------------------------------------------|-----------|
| <b>Executive summary .....</b>                                                                                       | <b>3</b>  |
| <b>1. Introduction .....</b>                                                                                         | <b>5</b>  |
| <b>2. Concept of data and distinction from similar concepts .....</b>                                                | <b>6</b>  |
| <b>3. Classification of data categories in platforms.....</b>                                                        | <b>8</b>  |
| <b>4. Map of policy issues.....</b>                                                                                  | <b>12</b> |
| Data sharing as a point of fairness in trading practices.....                                                        | 14        |
| Data sharing with third parties, stimulating innovation in aftermarkets.....                                         | 15        |
| Data and market power .....                                                                                          | 16        |
| Value of data .....                                                                                                  | 19        |
| Data ownership.....                                                                                                  | 20        |
| <b>5. Mapping of data-related practices in the Platforms' Economy: data collection, data access and sharing.....</b> | <b>25</b> |
| Data acquisition .....                                                                                               | 26        |
| Data sharing models .....                                                                                            | 29        |
| Tradeoffs in data .....                                                                                              | 39        |
| <b>6. Preliminary Conclusions .....</b>                                                                              | <b>40</b> |

# Executive summary



Online platforms are intermediaries in the digital economy that enable the exchange of goods, services or information between two or more parties<sup>1</sup>. They facilitate matching and make trade more efficient<sup>2</sup>. The mechanisms by which these digital intermediaries provide these efficiencies universally revolve around the use of technology that builds on data in one way or another. Such data is the subject of this report. Data generated through online platforms fosters innovation<sup>3</sup>. Data plays an increasingly important role in business intelligence, product development, and process optimization.

Data has become a new currency at times where many online services are provided for “free”, fuelled by the data provided by their users. Data is also the basis for competition and further innovation. While a number of national, EU and international reports clearly recognise the importance of data for the online platform economy, they rarely highlight the complexity and heterogeneity of data in the platform environment.

This report fills the gap by providing a structured overview of how data is generated, collected and used in the online platform economy. It maps out the diversity and heterogeneity of data-related practices and expands on what different types of data require a careful examination in order to better understand their importance for both the platforms and their users as well as the issues and challenges arising in their interactions.

The analysis first looks at different interpretations of the concept of data depending on whether it is observed from a philosophical/technical, economic or legal perspective. Considering concepts of data quality, the non-rival nature of data as well as issues of “data governance”, help understand different approaches to data depending on the context in which they are used.

There is a wide range of data-related practices and business models in the platform economy, reflecting the diversity of data, which any policy approach would need to consider. An overview of different data categories illustrates this variety, mapping platforms’ data practices,

---

<sup>1</sup> A conceptualization of platforms for legal and regulatory purposes in Rodríguez de las Heras Ballell, Teresa, “The Legal Anatomy of Electronic Platforms: A Prior Study to Assess the Need of a Law of Platforms in the EU”, 3 *The Italian Law Journal*, num. 1/3, 2017: 149-176, at 157.

<sup>2</sup> See, for instance Goldfarb, Avi, and Catherine Tucker. “Digital Economics”, *Journal of Economic Literature*, 2019, 57 (1): 3-43.

<sup>3</sup> EC (2017)

notably as regards their data collection and the access they allow to the data they hold as well as limits to such data access.

A number of policy issues arise in relation to how far the access to and possession of significant data sets by certain online platforms could distort the platform economy by giving rise to entry barriers, market power or preferential treatment of platforms' own goods and services. Emerging answers to these issues are not necessarily consensual, owing to the diversity of data and of the services as such in the online platform economy. The report examines different approaches to data usage, access and data sharing. It takes into account the value of data for platforms and businesses linked to its generation and processing on the one hand and to businesses' ability to access, analyse and use data on the other hand to improve their services.

The innovative capacities of platforms rely, in particular, on the richness of the data they collect, which allows them deep insights on how the markets function. In order to generate value across markets and stimulate innovation, however, different actors have an interest in getting access to data and opportunities to re-use it. The report looks into various incentives and models for data sharing, assessing whether there is room to promote data sharing beyond a mere remedy for market inefficiencies. It shows, on the other hand, that there may be legitimate reasons to restrict access to data, e.g. to protect trade secrets or a specific business model or even to prevent collusion.

The report concludes with a range of issues, which deserve, in the view of the authors, further policy attention and analysis. More research is required notably to address the lack of empirical evidence, in particular on data sharing practices and user needs. We would need more clarity on certain technical issues that might prevent or discourage data sharing. Finally, we would need to identify options for incentives as well as reasons for restraining the access to and sharing of data.

# 1. Introduction

A number of national, EU and international reports highlight the relevance of data for the online platform economy as well as for innovative, data-driven technologies such as AI. Yet, we have observed that these reports, somehow, lack detail as regards the use of the term “data”, disregarding the numerous facets that data can have, depending on how and by whom it is generated and used. This is why the Expert Group to the Observatory for the online platform economy has taken an academic and independent look at the **variety and complexity of data and its significance** for all different actors in the online platform economy. The purpose of this report is to inject a new point of departure in the discussions on data by providing a **structured overview of data-related practices in the online platform economy**. Specifically, we aim to summarise the present understanding of the importance of data for platforms’ business models, challenges associated with the socio-economic interactions that platforms intermediate and discuss the issue of access to data and data sharing.

The report comes timely at a moment where a number of government and private initiatives at national, EU and international level look into issues linked to data, calling for different approaches to regulate its collection, use, access and sharing, raising questions of competition, fairness, privacy and ownership. Against this background, the European Commission has just presented its data strategy for the next five years<sup>4</sup>. At the same time, it is launching a broad stakeholder consultation on the specific measures needed to ensure a leadership position for the EU in a data-driven economy, while respecting and promoting the European values.

The analysis should help moving away from general – and, implicitly, imprecise – statements on the need to share data – either as a voluntary initiative or potentially as binding commitments. It seeks in particular to provide nuance to the debate by elaborating on the types of data held by platforms and generated through their intermediation activities and identify the different interests at stake. The report provides a survey of data sharing practices and models and a discussion on alternative data sources and the value of data.

A multi-sided platform engages in a balancing act to internalize externalities across different sides. This also guides its policy to share data and data analytics with its business users. Does a platform have a socially optimal incentive to share data, in particular when data is an indispensable input for innovation activities of business users? How does the vertical integration of a platform affect its incentive to share data? The answers to those questions are important to guide policy makers. That is why, in this report, we try to investigate those issues both from theoretical and empirical points of view.

There can be three main reasons for which a platform does not share certain data with its business users. First, it may have a legitimate reason to protect its own business. For instance, some platforms, notably in the accommodation sector, usually do not share contact information of consumers with their business users, such as hotels, to prevent the latter from conducting transactions outside of the platform environment. Second, a complementor of a platform might

---

<sup>4</sup> [https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age\\_en](https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age_en)

grow to become a competitor of that platform. For instance, in the U.S. Microsoft case, Netscape, a complementor of Microsoft Windows, was expected to become a competitor of Windows. Last, vertical integration can induce a platform to restrict a rival app's access to data. For instance, Google seems to maintain exclusive access to Android APIs for some innovative functionalities (ARCEP, 2018).

In addition, there is currently a wider debate ongoing on e.g. platform-to-government data sharing<sup>5</sup>, on whether platforms should grant other actors – such as government authorities or potentially competing platforms – access to the data they hold, and whether businesses in general should share more data with each other. This report focuses primarily on business users' and other third-party access to data held or processed by platforms. Other angles of data sharing are treated in recent and ongoing work: for instance, the DG Competition Expert Group report of 2019<sup>6</sup> provides an overview of data sharing practices from a competition policy perspective. That report stresses that the area is one of emerging research, in particular in relation to the notion of so-called “data pools”, i.e. large scale coherent – but potentially unstructured – data assets held by private companies.

## 2. Concept of data and distinction from similar concepts

While many general statements are made about the importance of data in the platforms' economy, there is generally very little precision on what ‘data’ is, or which type of definition is used. Confusion often arises between volumes of data stored in data centres or transported through the physical network, or data as user-generated content disseminated through the platform, or, data as meta-information about users' interactions and transactions, online sales, etc. Consequently, technical, economic and legal definitions of data also need further context and precision.

From this perspective, data (an often numerical resource, a building block or preliminary stage of information) can be considered as distinct from information and knowledge.<sup>7</sup>

A well-established **philosophical** interpretation focuses on the “difference” that data express. Following this definition, “a datum is a putative fact regarding some difference or lack of

---

<sup>5</sup> <https://ec.europa.eu/digital-single-market/en/news/meetings-expert-group-business-government-data-sharing>.

<sup>6</sup> Jacques Crémer, Yves-Alexandre de Montjoye, Heike Schweitzer, *Competition policy for the Digital Era*, Publications office of the European Union, 2019.

<sup>7</sup> See e.g. for a technical perspective, the DIKW pyramid exposed in Rowley, Jennifer (2007). “The wisdom hierarchy: representations of the DIKW hierarchy”. *Journal of Information and Communication Science*. 33 (2): 163–180.; or Zins, Chaim (22 January 2007). [“Conceptual Approaches for Defining Data, Information, and Knowledge”](#) (PDF). *Journal of the American Society for Information Science and Technology*. 58 (4): 479–493.



uniformity within some context"<sup>8</sup>. The adjective “putative” is crucial here; it suggests that data are symbolic representations of the world that are neither objective nor free-standing givens. Instead, data have to be generated and made sense of and therefore reflect the context of their production. As the computer scientist Geoffrey C Bowker put it, "data are always already 'cooked' and never entirely 'raw'"<sup>9</sup>. This implies that the value of data can change over time and may be bound to specific contexts<sup>10</sup>.

Another school of thought focuses on the properties or affordance of data, understood as lending themselves to specific forms of analysis. Data are particulate or "corpuscular", not unlike sand, and as such aggregative<sup>11</sup>. They acquire their value as relational entries of data sets, which, in turn, may become components of other, larger data sets. Characteristic for digital data is that they can be "used, reused, copied, moved, and processed cheaply, without degradation, at very fast speeds" and that they "can drive economies of scale and scope"<sup>12</sup>. Due to digitalisation, the production mode of data is changing. To an increasing degree, data are today automatically generated by machines, programmes or sensors, not by human beings<sup>13</sup>. Even when data sets represent human beings or social actions, these may be involved only passively as a source rather than the producer of data<sup>14</sup>. Platforms generally instantiate fully digital infrastructures that support exchange and interaction through detailed user interfaces. Within these systems, user behaviour can then generate a wide variety of “unusual” data, including fine-grained logs of activities and temporal chains, taste and consumption patterns, and free-form expressions as text, audio, or video messages. As smartphones and Internet of Things (IoT) devices proliferate, movement through space and a broad array of sensor readings are added to fast-growing data pools, creating new opportunities for product and business development.

From an **economic** perspective, the use of data is non-rival: data can be used or copied without diminishing or exhausting them: “A single piece of data can be used in multiple algorithms and applications at the same time”, which can make it difficult ‘to establish the rights to use, exclude and transfer data’<sup>15</sup>. Digitalisation reinforces the non-rivalrous nature of data, with decreasing costs for copying, reusing and merging digital data sets<sup>16</sup>. The non-rivalry of data and reduced distribution cost implies greater latitude with regard to resource allocation, as the OECD<sup>17</sup> points out: "Non-rivalrous goods come with an additional degree of freedom with

---

<sup>8</sup> [Stanford Encyclopedia of Philosophy](#)

<sup>9</sup> cited after Gitelman 2013: 3

<sup>10</sup> for criteria regarding "data quality", see OECD (2015: 194)

<sup>11</sup> Gitelman (2013: 8)

<sup>12</sup> OECD (2019) [Data in the Digital Age](#)

<sup>13</sup> Duch-Brown et al (2017: 13)

<sup>14</sup> EPSC (2017: 5)

<sup>15</sup> [HM Treasury 2018: The economic value of data: discussion paper](#)

<sup>16</sup> Nestor Duch-Brown & Bertin Martens & Frank Mueller-Langer. "[The economics of ownership, access and trade in digital data](#)," JRC Working Papers on Digital Economy 2017-01, Joint Research Centre (Seville site), 2017.

<sup>17</sup> OECD (2015: 179-180)

respect to resource management. (...) social welfare is maximised not when the good is consumed solely by the person who values it the most, but when everyone who values it consumes it. Maximising access to the non-rivalrous good will in theory maximise social welfare, as every additional private benefit comes at no additional cost."

However, the general economic conceptualisation of data leaves a large leeway for a set of context-specific regimes. The non-rivalry of data allows for a variety of data governance models, which differ according to the distribution of control and access and their overall social welfare effects.

The **legal** approach to data or data-related practices focuses on rights relating to use, access, curation, control, stewardship or ownership. Specific properties of data affect the legal framework of "data governance"<sup>18</sup>. For the law to regulate data, it has to specify what it intends to protect<sup>19</sup>. For instance, the distinction between data syntax and semantics or that between machine- and human-produced data may be relevant for the appropriate legal framework. Other important properties concern the increasingly complex origins of and contributions to the production of data. Various "data stakeholders" are involved in the co-production of value and may therefore want to participate in the benefits<sup>20</sup>. However, traditional understandings of ownership "do not line up exactly" with the collaborative nature of data production in the platform economy and the de facto control over specific data sets<sup>21</sup>. More relevant than ownership titles to data might therefore be the control over "the means of analysis": While the 2015 OECD report still quotes A. Croll as saying that "The digital divide isn't about who owns data – it's about who can put that data to work"<sup>22</sup>, the debate has evolved in the meantime. Where platforms pretend that data sharing with other organisations would not help them as they would not know what to do and how to deal with the data they take a static and oversimplifying approach missing the development potential in many organisations. A stakeholder may have the means to analyse (or the general capacity to invest in getting to such means) huge datasets but when it has no realistic chance of getting access to relevant data it will not be able to perform such analysis (or start investing into obtaining the means). To conclude, the non-rivalry of data allows for a variety of data governance models, which differ according to the distribution of control and access and their overall social welfare effects<sup>23</sup>.

### 3. Classification of data categories in platforms

The wide variety of data-related practices in the platform economy reflects the great variety of data. Any policy approach to regulate data access or data sharing needs to consider this variety and heterogeneity of data. Therefore, the analysis of data-related practices in the platform

---

<sup>18</sup> see OECD (2015: 186)

<sup>19</sup> Drexl (2017: 263)

<sup>20</sup> OECD (2015: 195)

<sup>21</sup> OECD 2015: 195-7, with regard to patient data in the health sector; see also Drexl (2017: 260)

<sup>22</sup> Croll A. (2011) quoted in OECD (2015: 197)

<sup>23</sup> OECD (2015) Data-Driven Innovation: Big Data for Growth and Well-Being, (195-7)



economy requires some prior classification of data categories, not least for an accurate understanding of the business incentives and interests attached to each category of data, as well as to the legal implications of the respective categories.

A number of typologies of data are possible and several taxonomies have already been produced in different contexts. The aim of this report is not to contribute another classification of data, but to rely on existing typologies and assume an operational categorization of data for the purposes of the report and further evidence collection and inquiry. Thus, different taxonomies might be used side-by-side depending on the policy question that is being addressed.

A first categorisation, from a legal perspective, distinguishes between personal and non-personal data. Enshrined in the Charter of Fundamental Rights of the European Union (Article 11), the protection of personal data is further specified in the General Data Protection Regulation<sup>24</sup>, which defines personal data (Article 4(1) of that Regulation) as “any information relating to an identified or identifiable natural person”, e.g. name or address. In contrast, non-personal data does not pertain to an identified or identifiable individual, e.g. engine performance data, stock prices, etc. The borderline between the two categories is becoming increasingly blurred, in particular in an IoT context<sup>25</sup> while re-identification technologies are becoming more and more sophisticated<sup>26</sup>. Mixed data sets, including both personal and non-personal data bring additional complexities; the European Commission issued a set of guidelines aiming at adding further clarity as to the legal requirements concerning such data sets<sup>27</sup>. Ongoing research in areas such as differential privacy aims to reconcile data sharing and privacy concerns<sup>28</sup> while industry take-up needs to gain speed. Ultimately, it is clear that it is not enough to look at the data to understand the privacy and data protection risks that its processing generates: assessing the data environment is in fact the key to data governance best practice and to ensure effectiveness of processes, such as pseudonymisation and

---

<sup>24</sup> Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation); OJ L 119, 4.5.2016, p. 1

<sup>25</sup> For example, data generated by a smart thermostat about patterns of energy usage could be considered non-personal data. If, however, the data reveals behavioural patterns of identified or identifiable natural persons, the data could be categorized as personal data.

<sup>26</sup> See e.g. Rocher *et al.* “Estimating the success of re-identifications in incomplete datasets using generative models” Nature Communications 10 (2019) (showing that “the likelihood of a specific individual to have been correctly re-identified can be estimated with high accuracy even when the anonymized dataset is heavily incomplete”).

<sup>27</sup> <https://ec.europa.eu/digital-single-market/en/news/practical-guidance-businesses-how-process-mixed-datasets>

<sup>28</sup> See e.g. C. Dwork and A. Roth, “The Algorithmic Foundations of Differential Privacy”, *Foundations and Trends in Theoretical Computer Science* Vol. 9, Nos. 3–4 (2014) 211–407, DOI: 10.1561/04000000042;. Duchi, John; Jordan, Michael; Wainwright, Martin. (2013). “Local Privacy and Statistical Minimax”, in 54th Annual Symposium on Foundations of Computer Science.

anonymisation.<sup>29</sup> This consideration creates tensions with the traditional static approach to the scope of the data protection framework<sup>30</sup>.

A classification commonly used for personal data describes three categories according to its provenance: **volunteered data**, **observed data** and **inferred data**<sup>31</sup>. “Volunteered data”<sup>32</sup> is information that users actively share with the platform, e.g. when creating an account at an online marketplace or entering credit card details for online purchases. “Observed data” refers to activities of individuals, e.g. browsing history, location data from mobile devices. Finally, “inferred data” is information that a platform actively creates by aggregating, processing and analysing volunteered or observed data, e.g. customer risk profiles or analysed user sentiment. The collection of each category of data entails varying costs for the platform, with notable economies of scale.

From a privacy perspective, another important perspective comes from Article 7 of the Charter of Fundamental Rights of the European Union, guaranteeing privacy of communications. To the extent that platforms offer inter-personal communication services, or perform actions related to user terminal data, special provisions govern the use of such data sets.

In addition to these broad categorizations, some more granular categories have been suggested. A recent study distinguishes between six categories of data generated by online platform users, to facilitate an operational analysis of data sharing practices in the platform-to-business user relationship:<sup>33</sup>

- **Business identification details**, i.e. information on the business itself (company’s address, VAT number, country of operation).
- **User identification details**, i.e. information on customers/potential customers’ identity and profile (name, age, gender), contact details (email and delivery/home address), geographical provenance (IP address).
- **Data on individual transactions between businesses and customers**, i.e. the information generated through a specific transaction on the platform (good/service provided, price, payment method, communications between the business and the customer, reviews and ratings of the transaction, items viewed before/after the transaction and provenance on the Internet).

---

<sup>29</sup> Hence the importance of operating within a controlled environment.

<sup>30</sup> See e.g. Stalla-Bourdillon, S., & Knight, A., [Anonymous data v. personal data - a false debate: an EU perspective on anonymization, pseudonymization and personal data](#), *Wisconsin International Law Journal*, 2016, 34(2): 284-322. Some authors have thus argued that the distinction between personal data and non personal data should be abandoned. See e.g. [Purtova, N. \(2018\). The law of everything. Broad concept of personal data and future of EU data protection law. \*Law, Innovation and Technology\*, 10\(1\), 40-81.](#)

<sup>31</sup> World Economic Forum, *Rethinking Personal Data: A New Lens for Strengthening Trust*, May 2014; see also Cr  mer, Montjoye and Schweitzer (2019), Competition policy for the digital era, p. 24 et seq.

<sup>32</sup> The term “volunteered” is to be used with caution. In order to use a platform, users typically have no choice than to share their data. Data thus becomes an “involuntary currency”.

<sup>33</sup> VVA (2017) [Study on data in platform-to-business relations](#), p. 27.

- **Business performance**, i.e. information on all transactions taking place through the platform (number of product/service offered, prices and price changes, number of transactions through the platform, total value of sales, user traffic).
- **User behaviour**, i.e. data on customers’/potential customers’ behaviour on the platform (such as clicks, browsing history, other products or services purchased on the platform, geo-localisation, source of data (provenance), conversion rate).
- **Analyses of market trends/developments**, i.e. aggregated data and analysis of data collected by the platform and sometimes made (partly) available to business users ("data-as-a-service").

As this list indicates, data can be collected at different levels, e.g. for an individual transaction of a specific user (**individual-level data**), for the bundled transactions of a specific user (**bundled individual-level data**) or for a group of individual users (**aggregated-level data**).<sup>34</sup>

Another way of distinguishing different categories of data could be through a **sector-specific approach**, which takes into account that the value of data for business users very much depends on the specific business model and a given industry sector. For example, hospitality platform users value highly customer identification data (which is necessary to contact customers for promotional activities), whereas user behaviour data (which is necessary to target communication to individual customers) is of particular value for app stores and e-commerce platform users.<sup>35</sup>

In addition, depending on the policy objective pursued in the analysis, specific classification of different types of data need to be developed. For example, in the **relationship between data and platform market power**, specific categories of data may have a different impact on how platforms generate and increase their ability to retain users in their ecosystem, to consolidate their strategic position in a given market or to expand their activities to adjacent markets. For example, the exclusive access to customer identification details may be at the centre of platforms’ “intermediation power”<sup>36</sup>, as they mediate interactions between consumers and business users and thus aim to prevent their business users from switching and porting their customers’ details to other platforms or from concluding deals directly with their customers, outside the platforms’ environment (e.g. in the accommodation sector). In contrast, the ability to analyse market trends on the basis of aggregated transaction data can strengthen the market power of vertically integrated platforms, which could use such data for identifying new business opportunities<sup>37</sup> and foreclosing competitors in these markets.

---

<sup>34</sup> Crémer, Montjoye and Schweitzer (2019), *Competition policy for the digital era*, p. 25 et seq

<sup>35</sup> VVA (2017) [Study on data in platform-to-business relations](#), p. 35.

<sup>36</sup> Schweitzer et al. (2018), *Modernisierung der Missbrauchsaufsicht für marktmächtige Unternehmen*, p. 60 et seq.

<sup>37</sup> Zhu and Liu (2018), “Competing with complementors: An empirical look at Amazon.com”, 39 *Strategic Management Journal* 2618 at 2636; see also Belleflamme and Peitz (2019), “Managing Competition in a two-sided platform”, 28 *Journal of Economics & Management Strategy* 5 at 16. See also the competition

Further criteria to be taken into account for classifying data (and conditions for data sharing practices) should also include considerations of a technical nature – e.g. quality/structure, granularity, volume, frequency, real-time availability, format of the data – or the distinction between historic/real-time data, public/private data.

Finally, broad considerations about the source of data should also be accounted for<sup>38</sup>, in particular, where companies capitalise on network effects from one service to extend operations to more varied services, or when data is acquired from third parties. In particular, in the context of online advertising platforms, such complexity of data sources deserves further attention.

## 4. Map of policy issues

Data assets are instrumental to a well-functioning platform economy. At the same time, the same data assets also reinforce the most distinctive economic characteristics of online intermediation services through economies of scale, economies of scope, as well as through network effects. As such, access and possession of certain data assets can give rise to entry barriers, relative market power or even dominance, or differentiated treatment<sup>39</sup>.

These practices reflect the business strategy deployed by platforms to leverage and gain return on investment, effectively compete, retain users, and enhance their services, or create new services.

From the perspective of platforms' users, data flows are also bidirectional. Businesses have to provide certain data to be part of the platform's community, and as a condition to use a platform's services. Additionally, platforms observe, infer or predict<sup>40</sup> data from users' activities. In return, users expect to leverage their membership and exploit platform services to access valuable data of customers, market trends, competitors, forecasts, etc.

Platform businesses have pioneered many data-driven innovations. Successful platforms have built their value proposition on their ability to tailor and optimize their intermediation services, largely based on a variety of high value datasets they collect, creating positive feedback loops in building and reinforcing direct and indirect network effects. With the increased digitization of many aspects of our lives and economic transactions now going through online platforms, digital data offers a richness of opportunities but also brings adjacent risks<sup>41</sup>.

Amount, variety, availability/accessibility, quality, integrity and accuracy of data are critical factors in a highly data-intensive, data-driven, and data-dependant platform economy.

---

investigation the European Commission started against Amazon in July 2019:  
[https://ec.europa.eu/commission/presscorner/detail/en/IP\\_19\\_4291](https://ec.europa.eu/commission/presscorner/detail/en/IP_19_4291).

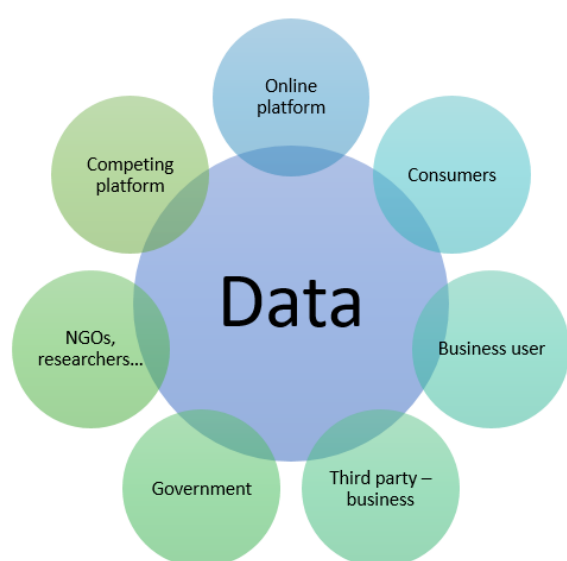
<sup>38</sup> Data provenance is an active area of research, which includes efforts for a standardised [provenance](#) framework

<sup>39</sup> Please see Progress Report of the Observatory on Differentiated Treatment.

<sup>40</sup> Or even predetermine the production of data through predetermination of future actions of actors.

<sup>41</sup> This report does not discuss the aspects related to personal data protection of such data acquisition. These aspects have been discussed elsewhere in depth.

Amongst other innovations, the richness and broadness of data collected by platforms are critical for the performance of innovative matching-capabilities (e.g. search or product ranking), for the performance of personalisation and recommender services, and for consolidating their competitive position in the market. In addition, platforms can have access to rich data on customer feedback, preferences between different products (for the whole market), trends, or consumers' behaviour on a scale not available to other business sectors, allowing them to continuously adapt and adjust their service. Thus, data are also a valuable source for platforms' adaptability and reactivity to markets' evolving needs. At the same time, platforms produce, process, and can make available valuable data for users (search results, comparison services, aggregated information, structured data, market trends, statistical information, predictions, shared data, etc.).



Data-related practices adopted by platforms are not inherently problematic *per se*.

While data is at the core of the platforms' business models, particular interests are emerging for a series of players in the broader platforms' ecosystem, from the interest of citizens to preserve and protect their fundamental rights, to other participants which the service intermediates (i.e. business users), to competing platforms, or opportunities for downstream innovation or, indeed, public policy concerns.

Data sharing can be seen as both, as a potential remedy for correcting market inefficiencies –

e.g. for addressing discriminatory practices or competition considerations in inter- and intra-platform dynamics – and as a policy objective in itself<sup>42</sup> – if we assume that further data sharing leads to wide use and data-driven innovation in downstream markets.

While data availability is certainly important, it is not sufficient in itself. The value of data (access) comes from its further innovative use. This requires the availability of the right tools, skills, technologies and infrastructures for data sharing.

In this context, a number of policy questions have emerged without necessarily generating wide-ranging consensus on the answers. This report will show that this is linked to the diversity of the role and nature of data in the online platform environment, in particular, in comparison with other sectors where data sharing has already been mandated or is being discussed (e.g.

---

<sup>42</sup> The objective of creating a common European data space is also specified in recital 35 of Regulation (EU) 2019/1150 on promoting fairness and transparency for business users of online intermediation services (P2B Regulation), OJ L186, p. 57, 11.7.2019

PSD2<sup>43</sup>, agriculture<sup>44</sup>, automotive, energy<sup>45</sup>). This is due to many factors such as the types of data at stake, the diversity of services offered in the platform economy as well as the different business models that exist. To exemplify this diversity, the report maps a number of policy challenges and provides an analytical summary of the status quo and known practices in the platform economy. It builds on preliminary work such as the report on “Competition policy for the digital era” and a literature review and formerly collected evidence sources. It will conclude with further questions and recommendations for collection of additional evidence necessary to formulate policy advice recognizing the heterogeneity of data issues in the online platform economy.

## Data sharing as a point of fairness in trading practices

Online platforms aggregate large amounts of personal and non-personal data. This is at the very core of their business models, of which excesses have been well documented and should require careful consideration<sup>46</sup>. Platforms’ ability to build and maintain a user base on both sides of the platform depends largely on the collection and retention of data. Most of this data collected by platforms is valuable for their business users as well as for third party businesses. It can help them to increase matching efficiencies, lower search costs, and contribute to innovation. However, instances have occurred where platforms refused to share data with their business users, even though in many cases the data is stemming from their business users or customers or was generated through their transactions and is then transformed into secondary data by the platform<sup>47</sup>. Data in platforms is also not exclusively coming from these actors, but platforms are also collecting data about consumers who are just browsing without buying from a specific business user.

At the starting point of the discussion around platform-held data lie two sets of claims by business users. First, they lack access to more individualised data about their own operations or own customers/consumers. Second, they require aggregated data on the entire market including also on competing business users’ operations and their respective

---

<sup>43</sup> Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC, OJ L 337, 23.12.2015, p. 35

<sup>44</sup> [https://www.copa-cogeca.eu/img/user/files/EU%20CODE/EU\\_Code\\_2018\\_web\\_version.pdf](https://www.copa-cogeca.eu/img/user/files/EU%20CODE/EU_Code_2018_web_version.pdf)

<sup>45</sup> cf. Article 23 on principles for the use of smart metering data generated by consumers of Directive (EU) 2019/944 on common rules for the internal market for electricity and amending Directive 2012/27/EU; OJ L 158 p. 125; 14.6.2019

<sup>46</sup> See e.g. Shoshana Zuboff, *The age of surveillance capitalism*, Profile Books; Main edition (31 Jan. 2019).

<sup>47</sup> See the Amazon competition investigation of July 2019:  
[https://ec.europa.eu/commission/presscorner/detail/en/IP\\_19\\_4291](https://ec.europa.eu/commission/presscorner/detail/en/IP_19_4291).



customers/consumers. 33% of the ‘heavy’ users of online platforms responding to a survey<sup>48</sup> stated that they had experienced data access related issues in their trading practices with online platforms.

While some platforms make data available to consumers, third-parties, or business users through APIs or data sharing portals, platforms can base their refusals for further sharing on data protection or business secrets considerations. Regulation (EU) 2019/1150 on promoting fairness and transparency for business users of online intermediation services (the so called “P2B Regulation”)<sup>49</sup> was conceived as a first step to ensure a more transparent and predictable trading environment for online business users. While it does not oblige platforms to share data with their business users it requires them, however, to be transparent about their data sharing practices with business users and third parties. As long as platforms provide such transparency, they may grant preferential data access to either their own businesses or businesses they control or to some of their business users and not to others.

To assess whether such transparency requirements are sufficient or whether additional measures are necessary, this report will provide an overview of different ways to look at data sharing, its benefits and drawbacks for the different parties involved.

### Data sharing with third parties, stimulating innovation in aftermarkets

Closely linked to the previous point on platforms' sharing of data with their own business users is the issue of data sharing with third party companies (e.g. other platforms, third party business users and data companies or data brokers who then trade data or data analytics), who could also make use of such data and potentially provide societal benefits by doing so.

#### **Examples:**

##### **The Finnish legislation on secondary use of health and social data<sup>50</sup>**

The Finnish government has introduced legislation to facilitate the secondary use of health and social data in compliance with EU Data Protection Legislation.

The purpose of the Act is to facilitate the effective and safe processing and access to the personal social and health data for steering, supervision, research, statistics and development in the health and social sector. The act’s intention is to stimulate innovation and development

---

<sup>48</sup> Ecorys, *Business-to-business trading practices in the online platform environment* (2017), study commissioned by the European Commission, available at <https://publications.europa.eu/en/publication-detail/-/publication/04c75b09-4b2b-11e7-aea8-01aa75ed71a1/language-en>

<sup>49</sup> OJ L186, p. 57, 11.7.2019

<sup>50</sup> <https://stm.fi/en/secondary-use-of-health-and-social-data>

activities by making aggregate data sets of individuals available to third parties, based on a digital permission system, integrated in a web service. While this is an example of a government-led innovation regulation not specifically targeted at platforms, it provides useful input for reflections on data sharing in other areas, too.

**The EU Code of conduct on agricultural data sharing**<sup>51</sup> by contractual agreement is an industry led example of data sharing enabling farmers in the EU to benefit from data driven strategies to add value to the agri-food chain and offer enhanced business opportunities. The rights regarding data produced on the farm or during farming operations are granted to (“owned by”) the farmers and may be used extensively by them. Data in the agricultural sector is generated during the various stages of agricultural production and related operations. This data is collected, transferred, processed and analysed. Data driven strategies can facilitate collective services, be helpful in negotiating fair contracts and facilitate the implementation of these contracts. The Code recognises the data originator’s right, whether they are a farmer or another party, to benefit from and or be compensated for the use of data created as part of their activity. It grants the data originator a leading role in controlling the access to and use of data from their business and to benefit from sharing their data.

Fully digital value chains (e.g. those integrating cloud and IoT services, or so-called “microservice architectures”) tend to be characterised by a greater amount of data sharing between parties. At the same time, data sharing between more traditional non-digital businesses and digital platforms is still an emerging area. The rise of B2B exchange platforms has led to business data accumulation by third parties, who use these data for analytics and value creation in a growing secondary data market.

Third party data sharing could enable the provision of input for ‘data pools’ established for a specific purpose (e.g. for AI training or charity or fundraising practices)<sup>52</sup> but also enable service providers to enter new markets. Reticence of platforms to share data, which could be relevant for business users to improve their own performance, e.g. on B2B IoT platforms, has led to a lack of trust in among the businesses using and depending on online platforms, which it is crucial to re-establish.

## Data and market power

---

<sup>51</sup> [https://www.copa-cogeca.eu/img/user/files/EU%20CODE/EU\\_Code\\_2018\\_web\\_version.pdf](https://www.copa-cogeca.eu/img/user/files/EU%20CODE/EU_Code_2018_web_version.pdf)

<sup>52</sup> This could take the form of a ‘data cooperative’ or ‘data trust’; models are only emerging.

The expert group report on "Competition policy for the digital era"<sup>53</sup> considered the competitive relevance of data for all actors in the online platform economy. It points out that "the competitiveness of firms will increasingly depend on timely access to relevant data and the ability to use that data to develop new, innovative applications and products".

Especially, data-driven AI innovations offer tremendous opportunities because algorithms based on deep learning have made important progress as exemplified in image and voice recognition, machine translation, health diagnostics and behavioural forecasting. Deep learning techniques, in particular, benefit from large amounts of data to derive their models from<sup>54</sup>. When analysing correlations, AI algorithms not only factor in strong features but also thousands of other weak features: peripheral data points that might appear unrelated to the outcome but contain some predictive power when combined across tens of millions of examples<sup>55</sup>.

Data reinforces each characteristic of platforms and thereby strengthens the tendency of market tipping and exposes the existence of economic power of platforms.

- a. Economies of scale: Most data are generated as a by-product of users' activities on online platforms. Once data are generated, the costs of (re)producing data are marginal. Hence, data reinforces economies of scale.
- b. Network effects: The larger the consumer base of a platform, the larger the amount of data the platform obtains. The larger the amount of data, the greater, in principle, the benefit (either quality increase or cost reduction) a platform obtains from the data. Therefore, data strengthens network effects<sup>56</sup>.
- c. Economies of scope: A platform with a large ecosystem operating in multiple markets can merge and combine different data sets to create a wide variety of data about its users, which allows the platform to gain in-depth insights about diverse practices and needs of its users. For this reason, data seems to be the strongest source for economies of scope.

For example, Ant Financial, a spinoff from Alibaba, represents one of the most spectacular data-driven developments of a platform ecosystem. Based on Alipay, the payment system of Alibaba, Ant Financial, has now nearly reached the double of the market capitalization of Goldman Sachs. Its strength lies in applying AI analytics to the enormous amount of data generated by the widespread use of mobile payments both for on-line and off-line transactions in China<sup>57</sup>.

---

<sup>53</sup> Crémer, Montjoye and Schweitzer (2019), *Competition policy for the digital era*

<sup>54</sup> We still need to nuance the statement that "big is better". This piece for example <https://arxiv.org/abs/1712.09471> suggests that even for unsupervised machine learning there is always an optimization goal underlying the analysis. Learning does not happen without knowing a target and information without a specific business need should always be discarded. More is not always better <https://analyticsindiamag.com/big-data-bias-variance-machine-learning-open-ai/>

<sup>55</sup> Lee (2018, p.111)

<sup>56</sup> Prufer and Schottmuller (2017) call it "data-driven network effects", see also Biglaiser et al. (2018)

<sup>57</sup> Zhu et al. (2017)

Both, network effects and economies of scope in data generate a number of important challenges.

- a. Positive feedback loop: An incumbent platform with a large market share naturally obtains a large amount of data from its users. The larger the amount of data, the higher the improvement in quality the platform can make, which in turn induces the platform to get an even larger market share, reinforcing the tendency for market tipping, which could risk leading to a monopoly.
- b. Data-driven entry barrier for actual or potential competitors: Data affects both, entrants' ability to expand or enter and incumbents' ability to defend and protect themselves. Without access to a significant amount of data, entrants may face difficulties to provide themselves a service of sufficiently good quality from entry to compete with an incumbent platform. In addition, incumbent platforms can use big data analytics to identify and neutralize potential threats of entry by taking defensive/offensive measures such as pre-emptive mergers and acquisitions, even if this does not always prevent the emergence of new platforms.<sup>58</sup>
- c. Data leverage within a platform ecosystem: A platform with a large ecosystem, i.e. which operates in multiple markets or sectors, can leverage its data advantage to enter an adjacent market served by specialized platforms<sup>59</sup>. When a new technology opens a new market, platforms with a large ecosystem may be the first to enter the market by leveraging their ecosystem and data power. For instance, Google, Apple and Amazon launched (or are expected to launch) their own cloud game platform to challenge the incumbents Microsoft and Sony.

In summary, data can reinforce tipping, create entry barriers and increase the economic power of platform ecosystems that benefit from strong network effects and economies of scope and scale in data. Monopolist platforms and powerful platform ecosystems, which i. a. thanks to their data related advantages (?) face no threat of entry from competitors, would have little incentive to innovate to increase consumer welfare. Hence, the potential benefit from data-driven innovations would remain under-exploited.

---

<sup>58</sup> See the description of how Facebook used a data-app, Onavo, to spot potential threats and buy or copy them in p.306 of Stucke, Maurice E., 2018. "Should we be concerned about data-opolies?" 2 *Georgetown Law Technology Review* 275

<https://georgetownlawtechreview.org/wp-content/uploads/2018/07/2.2-Stucke-pp-275-324.pdf>

On the role of data analytics in acquisitions see Lear, *Ex-post Assessment of Merger Control Decisions in Digital Markets*, 2019,

[https://www.learlab.com/wp-content/uploads/2019/06/CMA\\_past\\_digital\\_mergers\\_GOV.UK\\_version-1.pdf](https://www.learlab.com/wp-content/uploads/2019/06/CMA_past_digital_mergers_GOV.UK_version-1.pdf), p. 13, point 1.56; on acquisitions strategies see J. Furman, D. Coyle, A. Fletcher, D. McAuley, P. Marsden:

*Unlocking digital competition. Report of the Digital Competition Expert Panel*, London 2019, UK Government (Furman 2019), p.40; on pre-emptive acquisitions see Bourreau/Streel, *Digital Conglomerates and EU Competition Policy*, 2019, p. 21; further examples of usage of data for the purpose of acquisitions are reported in Khan, "The Amazon's Antitrust Paradox", *The Yale Law Journal* 2017, 710 (755, 780f.),

[https://www.yalelawjournal.org/pdf/e.710.Khan.805\\_zuvfyeh.pdf](https://www.yalelawjournal.org/pdf/e.710.Khan.805_zuvfyeh.pdf); regarding acquisitions by Facebook see Giulio/Scott Morton/Shapiro, *Antitrust and Innovation: Welcoming and Protecting Disruption*, 2019: 21, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3393911](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3393911).

<sup>59</sup> Stigler Center Report (2019)

The economic power of online platforms that act as gatekeepers is therefore a key challenge for the platform economy given the role of business data assets and the platforms' economic power to intermediate between businesses and consumers. It becomes even more critical in view of the gateway position platforms hold for a number of SMEs who need them to reach their consumers. It is therefore necessary, on the one hand, to look at competition *on* the market intermediated by platforms and the practices platforms apply to data sharing as well as on the other hand to look at competition between different platforms *for* a market in a specific sector and what role data sharing plays in this area. Another element related to competition and the possibility for businesses to multi-home or to switch service providers is linked to business' ability to benefit from data portability and interoperability. Therefore, it will be important to assess the effect that restrictions to portability and interoperability will have on innovation and consumer benefit and whether there are any “negative network effects” impacting at least some players in the online platform economy, and notably consumers.

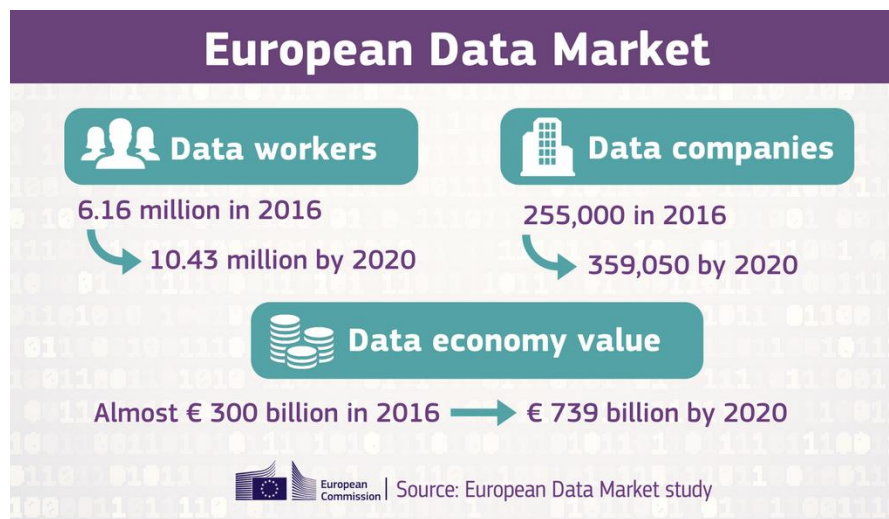
The power and attraction of a broader market for consumer data relevant for the entire economy, which is not limited to the classic advertisement industry, would also need to be considered. Platforms capture rich information about consumer preferences and behaviour that could sooner or later be useful for almost any consumer facing business. It could therefore be useful to analyse the specific status a few Big Tech companies have gained in capturing these data points and the structuring effect this has had on the entire consumer-facing economy, not just in their own and neighbouring markets but also in markets these companies may never want to enter. The huge and valuable data assets they have created cannot easily be replicated by other, notably emerging companies.

## Value of data

Consumers often ignore or underrate the value of the data they generate through their use of online platforms and their online interactions with businesses – often as a by-product of their online activities. At the same time, the large online platforms and search engines have become very skilled at understanding, evaluating and commercialising such information. Data driven network effects ensure that the better they perform, the more they can attract new business users and, again, more information through growing consumer interest. This causes a virtuous circle for those platforms that are already in or even leading the game, while it becomes much more difficult for new companies to enter and to catch up. To the extent that access to large data sets has become a requirement for data-driven innovation with the production and curation of data sets being a costly investment, dominant platforms are likely to benefit from their position over competitors and new market entries (Furman et al 2019: 4). Given the high level of concentration of digital platforms, the dynamics of data markets may reinforce the tendency towards monopolies in the data economy.

The European Data Market Study of 2017 has been measuring the size and trends of the EU data economy. According to its findings, the overall value of the data economy was almost reaching € 300 billion in 2016. According to the estimates of the data market monitoring tool, the value of the data economy in 2016 was worth nearly 2% of the European GDP. By 2020, the EU data economy is expected to increase to € 739 billion with an overall impact of 4% on

the EU DGP under a High Growth Scenario characterized by a stronger driving role for digital innovation and higher ICT investments.



The value of data for individual businesses is however difficult to assess and depends in particular on a business' ability to analyse and use data. Unlike tangible assets, data is non-rival, therefore data sharing does not reduce one's ability to use it or diminish its value, but sharing it may reduce a competitive advantage. Therefore, it is difficult to measure (and monetise) the value of data which might differ from one user to another or depend on the context and purpose data is intended for.

As mentioned earlier, we can distinguish between volunteered, observed or inferred data. One measure to determine the value of data could therefore be the cost of its production, which is naturally higher, where companies need to invest considerably to aggregate, process and analyse data to produce inferred data. The value of data can also be established by the existence of and competition with rival data sources and the willingness of businesses to pay a certain price to purchase it. Finally, the impact that certain data sets or, correspondingly, the lack of availability of good quality data can have on the economy and society can determine the value of data.

With the expected future growth of data on the one hand and its increasing economic and social value on the other, the question of appropriate governance structures arises. Future governance arrangements of the data economy will need to clarify rules and rights regarding control over, access to and use of data.

### Data 'ownership'



The data economy is still evolving as is our understanding of an appropriate legislative and regulatory regime for future transactions and usages of data. At present, control over data in the data economy is largely dominated by de facto technical control and contractual rights<sup>60</sup>. In the absence of any specific legal provisions, this leads to a de facto ownership, meaning that businesses or organizations in the position to collect data are able to exercise more or less exclusive control over it (albeit within the legal frameworks set by the GDPR and other relevant legislation).

Competition related concerns are just one of several issues regarding the emerging data economy. The conditions for realising its innovation potential present another challenge. Unlike personal data, which are traded by platforms and specialised information brokers<sup>61</sup>, industry or machine-generated data remain more difficult, to access outside their production context. Further research in this area will need to identify whether and how such data can be accessed and used and by whom. Although enormous quantities of data are being produced and stored in public and private sectors as a result of digitalization, and despite evidence for the demand for these data, barriers to data access still exist and risk lead to an under-utilisation of such data. Such under-utilisation occurs in situations "where the value of (secondary) data re-use for society (i.e. social and market value) is higher than the value (of primary data use) for the individual member of society (private value)"<sup>62</sup>.

Research on the need for regulation of the data economy, including viable approaches and their potential effects, is evolving at a high pace. While, in 2017, Kerber<sup>63</sup> still argued that it was not even clear whether the lack of access to and trade in data constituted a problem in need of public intervention or whether contractual arrangements among market players would be sufficient:<sup>64</sup>, research has evolved since, notably as regards IoT ecosystems<sup>65</sup>. Also different EU and international reports have advocated for effective ex-ante regulatory measures to complement competition law while at the same time reinforcing competition law enforcement<sup>66</sup>:

Part of the challenge that research on data governance models is facing can be explained by the non-rival nature of data. It is worth noting that the reduced cost of making data available implies greater economic and political latitude with regard to resource allocation. As the OECD points out<sup>67</sup>: "Non-rivalrous goods come with an additional degree of freedom with respect to resource management". However, notwithstanding its non-rivalry, third party use of data sets

---

<sup>60</sup> Duch-Brown, Martens and Mueller-Langer (2017)

<sup>61</sup> Crain (2018)

<sup>62</sup> OECD (2019)

<sup>63</sup> Kerber (2017:16)

<sup>64</sup> "a comprehensive analysis of market failure problems is still missing", Kerber (2017:11)

<sup>65</sup> [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3445422](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3445422)

<sup>66</sup> Cr  mer, Montjoye and Schweitzer (2019), *Competition policy for the digital era*, the [UK Furman Report](#), the [Stigler Centre report](#) or the [Australian Competition and Consumer Commission Report](#).

<sup>67</sup> OECD (2015: 179-180)

can be excluded. Data sets do not travel as easily as cultural goods such as video or audio files, as data structures and formats are generally not standardized across application sites.

To an increasing degree, data are today automatically generated by machines or programmes, not by human beings<sup>68</sup>. Even when data sets represent human beings or social actions, people may only be passively involved in data generation as a source rather than a producer<sup>69</sup>, and data production often involves more than one entity. In addition, data sets can be used in multiple contexts. All these aspects make an allocation of rights to access, use and transfer between producers and users a very complex task<sup>70</sup>.

Building upon the learnings of data sharing initiatives in different jurisdictions<sup>71</sup>, including initiatives funded by the European Commission such as the first EU-wide accelerator, Data Pitch<sup>72</sup>, an interdisciplinary debate around data governance is exploring various regulatory approaches for encouraging forms of sharing and granting access to data<sup>73</sup>. Zech<sup>74</sup> is one of the few scholars who investigates the possibilities of introducing a **producer right** to data. Legal ownership allocated to the "economically responsible operator of equipment that generates the data"<sup>75</sup>, he argues, would create incentives not only to produce but also to make available or reveal data<sup>76</sup>. As a result, new markets for reuse of data could evolve, which otherwise might not emerge at all. Moreover, rights to data property would increase transparency on available data, reduce uncertainty about data ownership and thereby potentially strengthen the efficiency of data markets<sup>77</sup>.

However, the matter of incentives for generating or granting third party access to data is subject to ongoing debates. Some scholars might argue that there is no evidence for a lack of incentives

---

<sup>68</sup> Duch-Brown et al (2017: 13)

<sup>69</sup> EPSC (2017: 5)

<sup>70</sup> HM Treasury (2018: 6); Datenethikkommission (2019: 104); Kerber (2017: 14)

<sup>71</sup> See e.g. GovLab. Data Collaboratives. Retrieved from: <http://datacollaboratives.org/>; Hub of All Things (HAT): <https://www.hubofallthings.com/>, Nesta

<sup>72</sup> <https://datapitch.eu/>. The lessons learned from the Data Pitch innovation programme are captured, in G. Thuermer, Johanna Walker & Elena Simperl, *Data Sharing Toolkit*, available <https://datapitch.eu/datasharingtoolkit/>. See also the European Data Incubator, <https://edincubator.eu/>.

<sup>73</sup> See recent literature on data trusts and data foundations, e.g. the work of the ODI <https://theodi.org/article/odi-data-trusts-report/> (e.g. Reed, C., BPE Solicitors & Pinsent Masons. (2019). Data trusts: legal and governance considerations); and the work of the Web Science Institute on data trust and data foundations, (Kieron O'Hara, [Data Trusts: Ethics, Architecture and Governance for Trustworthy Data Stewardship](#), 2019, Web Science Institute White Paper #1 and Sophie Stalla-Bourdillon, Alexis Wintour and Laura Carmichael, [Building Trust Through Data Foundations: A Call for a Data Governance Model to Support Trustworthy Data Sharing](#), Web Science Institute White Paper #2, available at <https://www.southampton.ac.uk/wsi/enterprise-and-impact/white-papers.page>). See also Delacroix, S. & Lawrence, N. (2019). Disturbing the 'One Size Fits All' Approach to Data Governance: Bottom-Up Data Trusts. *International Data Privacy Law*, DOI:<https://doi.org/10.1093/idpl/ipz014>.

<sup>74</sup> Zech (2016)

<sup>75</sup> Zech (2016: 10)

<sup>76</sup> Zech (2016: 11)

<sup>77</sup> Duch-Brown et al (2017)

for the production of data<sup>78</sup> and that the dominant regime of "de facto control of privately held data" seems to offer sufficient incentives<sup>79</sup>. Also, data ownership rights might not only reduce uncertainty but simultaneously create unintended incentives, among them incentives for the increased production of personal data and even the generation of artificial data. However, others take the view that the production of certain data, e.g. sensor data, rather results from an incentive to invest in the development and deployment of sensors which, again, depends on the potential to exploit such data or being rewarded otherwise (e.g. as a supplier).

The majority of scholars at present regards the introduction of data producer rights as a wrong and risky approach with various unknown effects. Since data constitutes not only the output of production but also the input for further innovative production, exclusive ownership rights in data might actually make innovation more costly, hamper market access and involuntarily support data monopolies. Another unknown effect concerns the interplay of a newly created data ownership with other relevant legal frameworks, among them the European data protection regulation, intellectual property law, the database directive and the protection of trade secrets<sup>80</sup>. Finally, as the OECD sums it up, traditional understandings of ownership "do not line up exactly" with the collaborative nature of data production in the platform economy, the distribution of potential rights to and factual control over specific data sets<sup>81</sup>.

The discussion about other regulatory options is in full swing. In its data strategy of February 2020<sup>82</sup>, the European Commission emphasises that, by 2030, the EU aims to "create a single European data space – a genuine single market for data, open to data from across the world – where personal as well as non-personal data, including sensitive business data, are secure and businesses also have easy access to an almost infinite amount of high-quality industrial data, boosting growth and creating value...", but also acknowledges that "[c]urrently there is not enough data available for innovative re-use, including for the development of artificial intelligence." The central underlying question is how data access and sharing can be encouraged even if "data access and sharing may benefit others more than it may benefit the data holder and controller, who may not be able to privatise all the benefits of data re-use. (...) The argument that follows is that if data are shared, free-riding users can 'consume the resources without paying an adequate contribution to investors, who in turn are unable to recoup their investments'"<sup>83</sup>.

Other barriers for sharing data consist in the perceived risk of violating data protection and/or intellectual property provisions but also the fear of losing control over the reuse of data. Given

---

<sup>78</sup> E.g. Kerber (2017)

<sup>79</sup> Kerber (2017: 9); Duch-Brown et al (2017: 21)

<sup>80</sup> Datenethikkommission (2019: 144); Hugenholtz (2017)

<sup>81</sup> OECD (2015: 195-7), with regard to patient data in the health sector; see also Drexl (2017: 260)

<sup>82</sup> [A European strategy for data](#), COM(2020) 66 final, 19.2.2020

<sup>83</sup> OECD (2019)

the specific properties of data, it can be difficult for data producers or providers to ensure compliance with agreed contractual terms.

Different approaches are considered to address the various barriers to data sharing. As a general principle, solutions will have to be context-specific. Access to public sector data requires different arrangements than access to privately held data, and personal data such as health data will need different levels of precaution than environmental information, for instance. As the OECD<sup>84</sup> notes, some member states are pioneering specific infrastructural provisions intended to mitigate the risks associated with data sharing. While most of these initiatives focus on access to public sector data, a few also target private sector data. Tools include guidelines and principles for contractual agreements and other kinds of data-based collaboration. Other forms of support are the certifying of data sharing platforms or the provision of a trusted third-party service<sup>85</sup>.

Another approach suggests the creation of a mandatory access right for specific users or types of data sets. Kerber<sup>86</sup> discusses specific access rights as an alternative to the concept of data ownership. Data that are of *general public interest* could benefit from such an access right. There are several national initiatives<sup>87</sup> under way to operationalise the concept of public interest data<sup>88</sup>. The European Commission pursues a slightly broader approach with a focus on access to private sector data for governments for the purpose of "a more targeted response to epidemics, better urban planning, improved road safety and traffic management, as well as better environmental protection, market monitoring or consumer protection". Relevant providers of such data could be digital platforms and social media, car manufacturers, retailers and telecom operators<sup>89</sup>.

## Other policy issues not addressed in this report

There are further policy concerns which, whilst relevant, will not form part of this report for several reasons, e.g. because they are out of the scope of this analysis or are already studied elsewhere such as:

---

<sup>84</sup> OECD (2019)

<sup>85</sup> for concrete examples, see OECD (2019)

<sup>86</sup> Kerber (2017: 16)

<sup>87</sup> The French Law for a digital Republic, for example, categorises "data of general interest" as "i) private-sector data from delegated public services such as utility or transportation services; ii) private-sector data that are essential for granting subsidies; and iii) private-sector data needed for national statistics"

<sup>88</sup> OECD (2019)

<sup>89</sup> European Commission (2018: 12)

- **Data governance for citizens**, notably the rights under the General Data Protection Regulation (GDPR)<sup>90</sup> allowing citizens to port their data to other services, without however granting them control over the data generated as a by-product as well as data security of citizens' data in an environment which is threatened in view of cybercrime or data breaches<sup>91</sup>.
- **Data sharing with competent public authorities or relevant third parties**, which would allow EU Member States' authorities to ensure regulatory compliance or to allow for judicial investigations. Another expert group<sup>92</sup> is presently studying this issue.
- **Data for good** to allow using and sharing data to benefit "not for profits" and non-governmental organizations by philanthropists' collectives, who want to help make society better through data (e.g. charities, cooperation with government initiatives, research etc.).

## 5. Mapping of data-related practices in the Platforms' Economy: data collection, data access and sharing

Platforms' innovative capacities rely, among others, on the richness of the data they collect. This is a unique feature of these digital services. Platforms, such as e-commerce platforms, are able to collect abundant data on their business users as well as on customers' previous purchases, reviews, preferences, search histories etc. - data, which allows platforms to continuously adapt and adjust their service<sup>93</sup>. The 2019 expert group report<sup>94</sup> stresses online marketplaces' unique position with regard to data: They "can observe in detail how the market functions". They can collect and analyse not only customer preferences but also the complementarity of products, how prices adjust etc.

This section intends to give a broad overview of data collection and data sharing practices in the platforms' environment. It also provides a general discussion on the use of data and the value-creation processes it entails.

---

<sup>90</sup> Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L 119, 4.5.2016, p. 1–88

<sup>91</sup> <https://www.theverge.com/2018/12/20/18150531/amazon-alexa-voice-recordings-wrong-user-gdpr-privacy-ai>

<sup>92</sup> <https://ec.europa.eu/digital-single-market/en/news/meetings-expert-group-business-government-data-sharing>

<sup>93</sup> See Martens, 2016 (p. 5), *Reinventing Capitalism in the Age of Big Data*, Viktor Mayer-Schonberger, and Thomas Ramge, 2018 (e.g. p. 77) and Zuboff's, *Age of Surveillance Capitalism*, 2019.

<sup>94</sup> Crémer, Montjoye and Schweitzer (2019), Competition policy for the digital era, p. 68.

This mapping and analysis exercise should be read as a non-exhaustive snapshot of the main types of behaviors, trying to illustrate a much vaster – and fast evolving – collection of practices.

## Data acquisition

In the analysis of the (big) data value chain, data collection is only one step in the process, followed by storage, analysis and usage<sup>95</sup>. This section focuses on data acquisition practices in the platform economy, in particular with regards to platforms' data collection practices as well as those of their business users; it does not elaborate on data collection practices for other participants in the broader ecosystem. It aims to identify issues for future consideration when developing further policy. While the extent and methods of data collection differ across the various types of online platforms, the below provides a non-exhaustive overview.

There are **several key issues** relating to data collection that could inform debates about the need and design of future policy measures, taking into account the various interests at stake: **(1) the cost of collection/acquisition**, **(2) incentives for collecting and/or acquiring data and competing interests**, and **(3) the availability and replicability of data**. If a certain set of data bears great value for several competitors but is not accessible and cannot be effectively replicated, there is potentially a stronger need to consider possible policy measures than in cases where data is readily available and is not costly to replicate. At the same time, the costs of data collection are sometimes high, and collection or production of data require considerable investment.

While data is inherently non-rivalrous, technical, legal and contractual restrictions may affect the non-rivalrous nature of data and raise the costs of data collection for a specific player. Technical barriers can for instance relate to data formats, data quality, and the quality of meta-data of data assets, or the distributed or encrypted nature of certain data assets – frequently necessary for security or data protection considerations. Protection of business secrets, data protection law, as well as privacy of electronic communications rules, form a relevant legal restraint for the collection and use of personal data. Contractual restrictions may include exclusivity clauses imposed on the transfer of data or limitations to the precise purpose of use of data sets<sup>96</sup>.

The cost (investment) needed to collect or produce the data is another relevant consideration. Here the **distinction between volunteered, observed and inferred data**, as discussed in the previous section, is relevant. As a first step, online platforms create the environment and maintain the service that leads to the emergence of “digital data traces” from their users, their

---

<sup>95</sup> See for instance: D.L. Rubinfeld & M.S. Gal, “Access barriers to big data”, *Arizona Law Review*, 2017: 349; E. Curry, “The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches”, in J.M Cavanillas & E. Curry, *New Horizons for a Data-Driven Economy - A Roadmap for Usage and Exploitation of Big Data in Europe*, Springer 2018, p. 32.

<sup>96</sup> See also the discussion in M. Bourreau, A. De Streel & I. Graef, *Big data and competition policy: Market power, personalised pricing and advertising*, CERRE project report February 2017, p. 30-31, available at <https://cerre.eu/publications/big-data-and-competition-policy>.



transactions and their online behaviour. While certain data classified as “volunteered” have to be provided as a prerequisite to use the services, platforms must invest in services’ attractiveness, which, inherently, prompts users to provide these data by using the service.

The tendency to collect more data is likely to continue for platforms to increase the quality and cost of their services, even though increasingly research is also focusing on extracting actionable insights from smaller, more anonymous, and/or distributed data-sets. Whereas the collection of volunteered data can be structured from the outset and, therefore, more easily and effectively processed, observed data may require a more intensive process as they are more heterogeneous.

Conversely, inferred data is information that platforms create themselves by analysing volunteered and observed data. Therefore, this type of data probably requires the most significant investment from platforms. While the detection of regularities and patterns in quantitative data may already require considerable investments in expertise and computational power, the analysis of textual, audio-visual, or sensor data can have significantly higher requirements.

Secondly, platforms need also to invest to produce **predictive data** (typically big data analytics) as well as to generate other own data. The amount, quality and variety of inferred and predictive data, as well as their ability to use the insights for business, process and product innovation, generally determine relative competitive advantages in the market of platforms. Unlike volunteered and observed data, inferred and predictive data depends upon platforms’ sophistication of analytics, the investment, and the ability to aggregate relevant data. Therefore, the size and quality of big datasets, analytical techniques and the financial capacity to exploit big datasets become mutually invigorating key assets in the platform economy. The incentives around data sharing are complex and can be related to the business strategies of the actors. On the one hand, it may be assumed that the more costly gaining and preserving these distinctive assets is, the greater the incentives for platforms to limit third party access to such data and enhance protection over them. On the other hand, business models that rely on growth through economies of scale, or scope, integration, or network effects more general may well choose to share data as part of a corporate growth strategy.

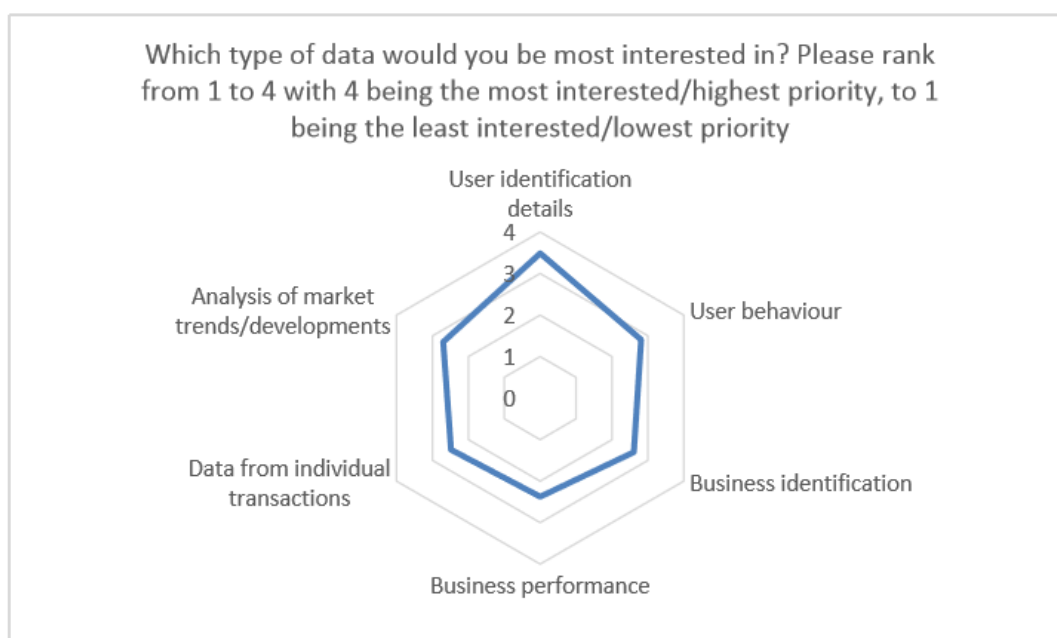
**Business users of online platforms** generally – but not always – cannot match the data-driven innovation capacity of online platforms. Empirical analysis<sup>97</sup> shows a great difference in the data savviness of different types of businesses – from large hotel chains, to very small hotel owners, from app developers to small shops selling products on market places, or, indeed, from large corporations monitoring and optimising their presence on search engines, to minor website owners. Business users can typically collect and curate (either independently or, sometimes, through interfaces and/or API access offered by online platforms) data resulting from their use of the platform, as well as transactional data from their exchanges with consumers. Such practices vary from one platform to another. In some cases, the particular use of data resulting from the use of the platform is contractually restricted. Some categories of

---

<sup>97</sup><https://publications.europa.eu/en/publication-detail/-/publication/4af6cec1-48fb-11e8-be1d-01aa75ed71a1/language-en>

business users seek greater access to customer data out of an interest in direct contact of consumers, disintermediating the transactions.

Further understanding is needed, on a cases-specific basis, for clarifying the data access needs, the current limitations in accessing data held by platforms – be it for lack of discoverability, technical challenges or refusals from platforms to grant access to the data sets - as well as the potential trade-offs between the interests of business users, of consumers (and the protection of personal data), and those of the online intermediaries. Moreover, with the entry into force of the “Platform-to-business” regulation, and in particular the provisions of Article 9 on access to data, platforms will have to provide enhanced transparency on their practices with regards to the data they collect, data collected by their business users, as well as sharing practices with third parties. It remains to be seen whether these transparency obligations will have a significant impact on practices; importantly, monitoring how platforms choose to comply is of paramount importance, both for the enforcement of the Regulation and for observing the practices and shifts in the platforms’ economy.



Source: VVA, Business user survey, 89 respondents to the question.

Finally, an important consideration stems from the **ability to collect data from alternative data sources**, allowing companies to limit the dependency on online platforms, or indeed triangulate data sources and build more valuable insights. This issue deserves further attention for each of the specific categories of data considered<sup>98</sup> in particular in assessing the dependencies data access may create as well as the costs linked to such approaches. Aggregated market insights, for instance, might be available elsewhere. Individual customer information details could potentially be collected at the point of transaction with individual consumers (whereas the pool of personal data of all consumers/users of the platform, could not).

---

<sup>98</sup> See chapter on categories of data, *supra*.

## Data sharing models

### Incentives for data sharing

Second step in the data value chain after data collection, data sharing practices are at the core of the data(-driven) economy and a pre-condition for generating value across markets through re-use of data. Data sharing practices need to be observed in the context of the **broader incentives for releasing or protecting data sets**.

Incentives for data sharing range from business strategies for testing innovation opportunities around the core business models (e.g. in giving real-time access to data to third party businesses which do not compete directly with the platform), or directly monetising data (e.g. in selling data sets to businesses or governments for specific applications), to technical considerations for interoperability of services, or as part of the value proposition and basic provision of the service (e.g. API access to the platform for business users of online marketplaces, aggregated reports on market trends and business intelligence shared with hotel owners for better pricing). Some companies are also incentivized to share data for reputational and public interest considerations (e.g. in sharing data with charities, public bodies or researchers).

Incentives for protecting data sets from being accessed by other entities are also quite diverse. Sometimes they include compliance with regulatory obligations, such as personal data protection or cybersecurity constraints, or protecting the service from DOS technical responses, or sometimes limiting access caused by crawler bots or to limit the cost of third-party data access.

Other times this relates to business considerations, restricting access to protect trade secrets to competitors, or protecting the core of the business model and breaking the network effects the value proposition builds on – i.e. avoid disintermediation of transactions between consumers and business users. Moreover, data sharing can, in specific cases, facilitate collusion or dominants' ability to leverage market power<sup>99</sup>.

Consequently, from a public policy perspective, both the encouragement of data sharing practices, and the restriction of data sharing are legitimate objectives under specific circumstances.

### Modalities for data sharing – main considerations

Further, we discuss, in this section, various **models of data sharing**. First, in terms of modalities of data sharing, a number of useful distinctions could be made, as described below, and illustrated with a number of examples. These are not necessarily offered as hard

---

<sup>99</sup> J. Crémer, Y.-A. de Montjoye, H. Schweitzer, *Competition Policy for the Digital Era: Final Report*, 2019: at 68, 85, 96.

recommendations for best practices, but rather meant either to describe established practices, or point to experimental efforts challenging the status quo.

*One-off data sharing versus real-time, continuous access.* Both models are legitimate and appropriate in different situations. Real-time ‘data hose’ established through a performance APIs is typically a precondition for valuable re-use of data in tech-savvy environments (e.g. building a service ‘on top’ of the data access itself, or inspecting and researching socio-technical phenomena in a data-right environment<sup>100</sup>), but less so in cases the recipient does not have the data analytics skills and the infrastructure to make use of the opportunity.

Where data access is legally mandated, different regimes exist. Opinions are divided over the question whether the data portability right enshrined in Article 20 of the General Data Protection Regulation offers a one-off access to data only<sup>101102</sup> or whether it can also be interpreted to provide for continuous data access, in particular in sectors where specific data access regimes have already been adopted (such as the access-to-account obligation in the Payment Services Directive which establishes real-time continuous data access)<sup>103</sup>.

*Data sharing for free or against remuneration* – there is a range of models under which data is shared, intricately related to the incentives for data sharing. Data can be shared for free, as an ancillary service or as part of a contractual obligation, or it can make the object of a business transaction in itself, for instance under fair, reasonable and non-discriminatory conditions (FRAND).

#### **Example: Twitter API**

The Twitter social network provides access to several types of APIs facilitating access to different functionalities of the platform<sup>104</sup>. For accessing real-time data (tweets and metadata), three different plans are available, ranging from free of charge (limited to a small sample of tweets), to premium and enterprise versions of the API, against different pricing schemes.

In 2015, Twitter announced that it was ending third-party agreements for the resale of Twitter’s so-called “firehose” data, referring to the full stream of tweets available from the service.

---

<sup>100</sup> See, for example, discussions on API access to ad data for scrutinising disinformation concerns <https://blog.mozilla.org/blog/2019/03/27/facebook-and-google-this-is-what-an-effective-ad-archive-api-looks-like/>

<sup>101</sup> As suggested in Crémer/De Montjoye/Schweitzer (2019), *Competition policy for the digital era*, p. 81-82.

<sup>102</sup> Graef/Husovec/Purtova, “Data portability and data control”, *German Law Journal* 2018: 1959; referring also Schweitzer, *Datenzugang in der Datenökonomie: Eckpfeiler einer neuen Informationsordnung*, GRUR 2019, 569 (576) and Fn 42.

<sup>103</sup> Graef/Husovec/Van den Boom, “Spill-overs in data governance: Uncovering the uneasy relationship between the GDPR’s right to data portability and EU sector-specific data access regimes”, *Journal of European Consumer and Market Law* 2020 (forthcoming).

<sup>104</sup> <https://developer.twitter.com/en/products/products-overview>

Third-party data analytics services who had built business models on the basis of access to Twitter's firehose were significantly affected by the change, highlighting the complex nature of dependencies in the API economy<sup>105</sup>. As a particularity, Twitter provides a representative sample of tweets in real-time, which allows for statistical processing without full access, constituting an interesting model in its own right.

*Voluntary data sharing vs. compulsory access.* While most of the data sharing practices today stem from voluntary agreements and market dynamics, some regulatory models also address specific issues. Self- and co-regulatory practices apply in a variety of sectors<sup>106</sup>, and the creation of standard contracts by government or setting up other initiatives such as the “data sharing coalition” of the Dutch Ministry of Economics<sup>107</sup>. Mandatory data access conditions are also set in a number of legal acts from the transport to the financial services sector. Data access, or indeed, conditions for portability and interoperability are also explored in the specifics of competition remedies.<sup>108</sup>

*Data sharing with direct or indirect competitors* is the exception, rather than the rule, notably as compliance with competition needs to be ensured when competitors share commercially sensitive data. The scenario for data sharing with competitors as a remedy for dominance was long discussed in the literature - e.g. around search engine data, with inconclusive results as to the effectiveness of the remedy. In many cases, data is publicly available *de facto*, but companies limit its “capture” and use by third parties through the terms and conditions of their websites and apps. (Google to share search data with a new entrant that wishes to set up its own search engine – this is a hypothetical example) or indirect competitors (Twitter data used for data analytics services or US HiQ/LinkedIn case<sup>109</sup>).

**Example: hiQ Labs vs. LinkedIn (US)**

hiQ Labs initiated a lawsuit against LinkedIn when the latter issued a cease and desist order against hiQ Labs for scraping data from the publicly available profiles listed by LinkedIn in order to further analyse the data and provide its HR intelligence services. The Californian court decided<sup>110</sup> that the practice of scraping publicly available data did not infringe the American Computer Fraud and Abuse Act of 1986 (CFAA).

---

<sup>105</sup> See e.g. <https://www.programmableweb.com/news/how-datasift-survived-twitters-merciless-business-behavior-api-economy/native-case-study/2018/12/11>

<sup>106</sup> E.g. for agricultural data, see the EU Code of Conduct on agricultural data sharing by contractual agreement, [https://copa-cogeca.eu/img/user/files/EU%20CODE/EU\\_Code\\_2018\\_web\\_version.pdf](https://copa-cogeca.eu/img/user/files/EU%20CODE/EU_Code_2018_web_version.pdf).

<sup>107</sup> Dutch vision on data sharing between business at <https://www.government.nl/documents/reports/2019/02/01/dutch-vision-on-data-sharing-between-businesses>.

<sup>108</sup> See, more in detail, J. Crémer, Y.-A. de Montjoye, H. Schweitzer: *Competition Policy for the Digital Era: Final Report*, 2019, p. 83,84.

<sup>109</sup> <https://www.tripwire.com/state-of-security/featured/hiq-v-linkedin-controls-publicly-available-data/>.

<sup>110</sup> <http://cdn.ca9.uscourts.gov/datastore/opinions/2019/09/09/17-16783.pdf>

*Data portability and interoperability* discussions are closely related to data sharing objectives but follow a slightly different policy concern. Data portability is generally understood as the ability to port (transfer) data from one service to another. Interoperability refers to the technical features of the data infrastructure and standardisation of data, allowing different technical systems to interact. Both measures follow voluntary cooperation schemes as well as, in some cases, regulatory provisions, as in the General Data Protection Regulation, the Payment Services Directive, or the Digital Content Directive. At the same time, private initiatives for facilitating interoperability of services are also emerging, with yet unclear specifications for the governance models or the inclusiveness of the services, which could benefit from the portability features. Large companies are initiating such efforts amongst themselves, while privacy advocates are proposing initiatives structured around “MyData” models<sup>111</sup>, with an aim to empower data subjects to make use of their portability right on personal data<sup>112</sup>. Interoperability and standardization can more generally be seen as key facilitators for the sharing of data, and interoperability in particular is a prerequisite for the real-time, continuous access to data mentioned above.

**Example: The Data Transfer Project (DTP)<sup>113</sup>**

The Data Transfer Project is an initiative shared by five major digital services – Google, Apple, Facebook, Microsoft and Twitter – aiming at developing an interoperable infrastructure across their services, “with open-source code that can connect any two online service providers, enabling a seamless, direct, user initiated portability of data between the two platforms”. The project is not yet operational, though some documentation is available under an open access regime. Use cases announced by the project include facilitating transfer of data to a third-party service, switching a service, backing up data. The actual practical use, take-up and implications will provide further insights into such industry-led projects on data sharing.

**Example: Social Web Protocols<sup>114</sup>**

The Social Web protocols, developed under a W3C working group, proposed a set of protocols for social web interactions, including data syntax protocols, client-side APIs and web protocol, as well as an underlying linked data platform. This alternative model was concluded in 2018,

---

<sup>111</sup> Antti Poikola, Kai Kuikkaniemi & Harri Honko, *MyData – A Nordic Model for human-centered personal data management and processing*, 2014.

<sup>112</sup> See, for example, the P3C Consortium and the Privacy Tech White Book <https://www.privacytech.fr/livre-blanc-privacytech.pdf>

<sup>113</sup> <https://datatransferproject.dev/>

<sup>114</sup> [www.w3.org/TR/social-web-protocols/](http://www.w3.org/TR/social-web-protocols/)



and the community behind it continues experimentation under a Social Web Incubator Community Group<sup>115</sup>.

*Data sharing when sensitive data is concerned.* One of the outstanding challenges in the data (sharing) economy is linked to the potential trade-offs of data protection and security when data is released. While some efforts are ongoing to further methodologies of pseudonymization of data sets and differential privacy, research also points to risks of re-identification even based on features unsuspected to identify a person<sup>116</sup>. Further discussions emerge, however, on different models managing access to data, where the actual transfer of data is not necessary in itself, whereas the computation and analysis can take place remotely. Some technical and governance models are emerging that offer promising avenues for facilitating the value extraction from data sets, while mitigating potential risks<sup>117</sup>.

**Example: The OPAL project (Open Algorithm)<sup>118</sup>**

The OPAL project is a not-for-profit initiative of the Data-Pop Alliance, Imperial College London and MIT Media Lab, Orange and the World Economic Forum, aiming at facilitating the use of private sector data for the public good. It is currently focused on telecommunications data, with two experiments in Colombia and Senegal, and deeply linked to the use of privately held (sensitive) data in support of the UN Sustainable Development Goals. The project is developed on two dimensions: (1) a technology infrastructure, including open algorithms querying pseudonymized data sets stored on the server of the company holding the data and sharing aggregate statistics; (2) a governance framework including oversight bodies and participatory design and testing of the features. While the project does not, currently, foster platforms' held data, it presents a pioneering model for the technical set-up and the governance of data sharing and use.

*Data sharing through data markets.* Data sharing can be a by-product of a commercial exchange, but it can also be the main purpose of the interaction. Data marketplaces are emerging in a number of fields as resources for identifying and potentially commercialising valuable data sets and exchanges. The data 'offer' is generally brought by so-called data brokers or aggregator services, and different models of data marketplaces are emerging.

---

<sup>115</sup> <https://www.w3.org/wiki/SocialCG>

<sup>116</sup> E.g. location data, and models can also remember training data, see <https://arxiv.org/pdf/1709.07886.pdf>

<sup>117</sup> INSEE offers this service, see [www.casd.eu](http://www.casd.eu)

<sup>118</sup> <https://www.opalproject.org/>

**Example: DAWEX<sup>119</sup>**

DAWEX<sup>120</sup> Data Exchange Platform is a global data marketplace that enables actors to share, monetize, and acquire data without intermediaries. It is designed for all types of private companies and public-sector organizations, which can coordinate the flow of data by sourcing and exchanging data securely and in compliance with applicable regulations, ensuring the integrity of license agreements thanks to block-chain technology to manage the security, safety and transparency of exchanges.

Platforms' data sharing with businesses

In principle, businesses can gain access to data collected and/or gathered by online services through multiple channels. The following table provides a non-exhaustive overview of the key types of data access models. The initial five data access models (i.-v.) are largely based on two reports – [5, p. 5] and [6, pp. 60-65] – published in 2018.<sup>121</sup> We have added data innovation programmes (vi.) and some additional examples. Note that these examples are relatively broad in scope and might need to be reviewed to better stick to one population of online services.

---

<sup>119</sup> Dawex' origins are in France. It won the Digital Innovation Competition in 2015 and received funding under the European Union's Horizon 2020 research and innovation programme in 2018

<sup>120</sup> <https://www.dawex.com/en/>

<sup>121</sup> Note that these two reports focus on B2B data sharing – we have used P2B data sharing interchangeably with this term.

| Table 2. An overview of the key types data access models for P2B data sharing |                                                                                                                                              |                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                    |
|-------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                               | Type of P2B data sharing model                                                                                                               | Approach to data sharing                                                                                                                                                                                                                                                     | Potential key motivations for this approach                                                                                                                                                                                                                                                                                                                                                                                                                           | Some examples                                                                                                                                      |
| i.                                                                            | Open data model.<br><br>Also referred to as “ <i>an open data approach</i> ” <sup>122</sup> , and “ <i>open data policy</i> ” <sup>123</sup> | An online platform makes data accessible to a wide-range of re-users with (a) “ <i>as few restrictions as possible</i> ”, and (b) “ <i>no or very limited remuneration</i> ” <sup>124</sup> (e.g. a small payment is required to cover re-formatting costs <sup>125</sup> ). | An online platform (a) may provide open data where it has “ <i>a strong interest</i> ” in data re-usage <sup>126</sup> – e.g. for innovation purposes <sup>127</sup> . Such a strong interest in re-usage may be rooted in “ <i>data philanthropy</i> ”, “ <i>corporate social responsibility</i> ” and “ <i>strategic interests</i> ” <sup>128</sup> . An online platform (b) can be <b>legally required</b> to make certain data openly accessible <sup>129</sup> . | E.g. Google Trends <sup>130</sup> , Google Books Ngram Viewer, Google Finance, Amazon Web Services Open Data, Million Song Dataset. <sup>131</sup> |

<sup>122</sup> European Commission (2018): Towards a common European Data Space p. 5

<sup>123</sup> E. Scaria, A. Berghmans, M. Pont, C. Arnaut and S. Leconte, "Study on data sharing between companies in Europe", p. 64. For further background information on open data, the Open Data Institute (ODI) examines how three big business create value with open data.

<sup>124</sup> European Commission (2018): Towards a common European Data Space p. 5

<sup>125</sup> E. Scaria et al., p. 64

<sup>126</sup> European Commission (2018): Towards a common European Data Space p. 5

<sup>127</sup> E. Scaria et al., p. 64

<sup>128</sup> European Commission (2017), p. 14

<sup>129</sup> E. Scaria et al., p. 64

<sup>130</sup> European Commission (2017)

<sup>131</sup> For further examples of open datasets see B. Marr (2016)

|      |                                                                                                                                                                                                                                                                                                                   |                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                |                                                                                                                                                  |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| ii.  | Data monetisation model <sup>132</sup> .                                                                                                                                                                                                                                                                          | An online platform makes data accessible via a “ <i>unilateral approach</i> ” – e.g. through a particular licence <sup>133</sup> .                                                                          | An online platform wants to (a) <b>generate “revenue”</b> through data sharing and, in some cases, add further value via (b) “ <i>the provision of services</i> ” <sup>134</sup> .                                                                                                             | E.g. “Telefónica [...] provides access to anonymised or aggregated insights derived from the data that the company holds [...]” <sup>135</sup> . |
| iii. | Data marketplace model. <sup>136</sup><br>Also referred to as “ <i>data monetisation on a data marketplace</i> ” <sup>137</sup> – this could be considered as a sub-category of data monetisation. Furthermore, open data marketplaces (e.g. G. Smith e.a. (2016), Deloitte (2012)) could be considered as a sub- | An online platform uses a trusted third party intermediary – i.e. a data marketplace – for data transactions <sup>138</sup> . E.g. data access may be provided through bilateral contracts <sup>139</sup> . | An online platform (a) is unsure about the “ <i>potential re-users for their data</i> ” – and is able to <b>test the market</b> through “ <i>one-off data monetisation efforts</i> ” <sup>140</sup> . The company wants to (b) <b>generate “revenue”</b> through data sharing <sup>141</sup> . | E.g. Dawex, Data Intelligence Hub.                                                                                                               |

<sup>132</sup> E. Scaria et al., p. 61

<sup>133</sup> E. Scaria et al., p. 61

<sup>134</sup> E. Scaria et al., p. 61

<sup>135</sup> E. Scaria et al., p. 61

<sup>136</sup> E. Scaria et al., p. 62. For further background information on data marketplaces see Florian Stahl et al.

<sup>137</sup> European Commission (2018): Towards a common European Data Space, p. 5

<sup>138</sup> European Commission (2018): Towards a common European Data Space, p. 5; E. Scaria et al., p. 62

<sup>139</sup> European Commission (2018): Towards a common European Data Space, p. 5

<sup>140</sup> European Commission (2018): Towards a common European Data Space, p. 5

<sup>141</sup> E. Scaria et al., p. 62

|     |                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                       |                                                                                                                        |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------|
|     | category of open data models.                                                                                                                                                                                                                           |                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                       |                                                                                                                        |
| iv. | <p>Exclusive data platforms.</p> <p>Also referred to as “data exchange in a closed platform”<sup>142</sup>, “industrial data platforms”<sup>143</sup>, “direct data exchange between two companies”<sup>144</sup> and “data pooling”<sup>145</sup>.</p> | <p>An online platform shares data via a closed platform (e.g. independently, via a third party intermediary)<sup>146</sup>. For instance, a restricted group of users may voluntarily join an industrial data platform<sup>147</sup>.</p> | <p>An online platform may want to: (a) obtain “monetary remuneration”; (b) offer “value-added services”; (c) create more “stable data partnerships”; and/or (d) retain greater “control” over the re-usage of data<sup>148</sup>.</p> | <p>E.g. Industrial data platforms: “Airbus and MAN created Skywise and RIO platform (respectively)”<sup>149</sup>.</p> |

---

<sup>142</sup> European Commission (2018): Towards a common European Data Space, p. 5

<sup>143</sup> E. Scaria et al., p. 62

<sup>144</sup> H. Richter and P. R. Slowinski (2019)

<sup>145</sup> H. Richter and P. R. Slowinski (2019)

<sup>146</sup> European Commission (2018): Towards a common European Data Space, p. 5

<sup>147</sup> E. Scaria et al., p. 62

<sup>148</sup> European Commission (2018): Towards a common European Data Space, p. 5

<sup>149</sup> E. Scaria et al., p. 62

|    |                                     |                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                            |                                                                   |
|----|-------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------|
| v. | Technical enablers <sup>150</sup> . | A technical enabler is a third party company that facilitates B2B data sharing between other companies through a “technical solution” <sup>151</sup> . Technical enablers are not data marketplaces or exclusive data platforms; revenue is generated through the “ <i>set-up, implementation and maintenance of their solutions</i> ” <sup>152</sup> . | An online platform may work with a technical enabler in order to (a) “ <i>exchange data within a particular community in a more agile way</i> ”; and (b) utilise and customise existing tools for P2B data sharing rather than “ <i>investing resources in developing something new</i> ” <sup>153</sup> . | E.g. “API-AGRO, DKE-Data, Nallian and Sensative” <sup>154</sup> . |
|----|-------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------|

---

<sup>150</sup> E. Scaria et al., p. 63

<sup>151</sup> E. Scaria et al., p. 63

<sup>152</sup> E. Scaria et al., p. 63

<sup>153</sup> E. Scaria et al., p. 63

<sup>154</sup> E. Scaria et al., p. 63



## Tradeoffs in data

Over the last 15 years, the digital transformation has brought about an enormous growth of data, notably on on-line platforms, new means of analysing big data sets and new opportunities for realising commercial and public value. Data enable data-based innovations across the public and private sectors, allowing new business models and markets to emerge, which are specifically designed for generating, collecting and commodifying data. Digital platforms, which collect data about their users and the interactions with and among them, are examples for this development. Some of the prominent platforms play a pioneering role in the development of data-driven technologies and applications. This also includes the use of machine learning processes in the analysis of data and the creation of AI-based services<sup>155</sup>.

An [EC Study on B2B data sharing of 2018](#) found that a considerable proportion of companies which are not yet engaged in B2B data sharing and re-use, expect to start sharing and re-using data in the next five years. While the European data economy is still at its very beginning, companies already collect, share and re-use data among them to innovate, develop new products and services and/or new business models. They have recognised the potential and benefits of these activities and the new opportunities for increase network effects, which help improving their services as well as their relationship with their customers and generating additional revenues.

The study observed, however, that companies that engage in B2B data sharing do not necessarily grant access to their complete datasets. The proportion of data shared by companies usually depends on their business strategy. As a result, data may be shared against payment, or companies preserve a certain discretionary autonomy with whom they want to share what data and under which conditions.

Not investing enough in access to relevant data can have negative consequences for companies such as missed business opportunities to improve relations with their customers, but also to innovate, scale up or streamline administrative processes.

There are a number of different actors in the online ecosystem, which is cross-border and inter-sectorial by nature. These different actors obviously have different interests in relation to data sharing. The EC study distinguishes notably between data suppliers and data users. “*Data suppliers* make datasets available to interested data users in compliance with relevant laws. This means that they hold “ownership”-type of rights over data and/or have obtained legal permission from private individuals to share their data. *Data users* access data from data suppliers following the conditions set up and agreed on between them. Companies may play a dual role in this ecosystem, i.e. they are both data suppliers and data users. In certain cases, companies decide to engage in strategic partnerships to exchange data among them. *Industrial data platforms* formalise and operationalise

---

<sup>155</sup> ACCC, 2019: 11.

these alliances, which can encompass different business sectors, or be limited to a specific industry.”

In addition to the opportunities that data sharing between online platforms and their business users undoubtedly offers, there are also responsibilities, such as GDPR compliance, and other legal, technical and financial burdens on both platforms and their business users. Moreover, questions of proportionality, purpose limitation and non-discrimination arise as well as those of the rights of the data subjects.

## 6. Preliminary Conclusions

This report highlights the diversity of what is commonly subsumed under the heading “data”. This is indeed not a homogenous, clearly delineated concept, and the design of any public policy should reflect this. This means in particular that an all-embracing, horizontal policy approach, disregarding the heterogeneity of data in the platform economy and the variety of business-specific practices, would not be appropriate

More research is needed to better understand the main long-term incentives as well as the potential and limits of voluntary and mandatory tools for platforms to collect, analyse and share data. Further research is also needed to assess the viability of different, potentially competing models, which may be based on free access to and use of data for (certain) third parties, (limited) data-sharing or data ownership. Additional work also needs to be devoted to the assessment of an overall positive effect of data sharing and on weighing the risks of the latter against potential benefits for competition, value creation and innovation.

In this context, it is important to take into account privacy and data protection rules and also incentives for platforms to establish and scale up. This is all the more relevant for newly emerging platforms, whose main assets may be the data they collect. Horizontal data sharing obligations could put at risk those new players’ existence.

As a consequence, here are some specific areas which could be focal in the next phase of our work on data and are also consistent with the recommendations of the DG Competition expert group report.

1. Empirical evidence (e.g. on data sharing practices in a range of intermediaries, varying needs depending on the users and their specific purposes or issues relating to the indispensability and replicability of data in the platform economy).
2. Technical issues (e.g. overview of the ecosystem of APIs and related emerging technology interfaces for data-sharing).
3. Solutions covering incentives as well as constraints for data sharing (e.g. discriminatory access to data, exclusivity arrangements, examination of opportunities, risks and benefits of a “reversal of the burden of proof” for refusals to share).

## Bibliography

"Amazon Web Services," [Online]. Available: <https://aws.amazon.com/opendata/>

[Australian Competition & Consumer Commission](#) (ACCC) (2019): Digital Platforms Inquiry. Final report. <https://www.accc.gov.au/publications/digital-platforms-inquiry-final-report>

O. Batura, N. van Gorp and P. Larouche , "Online Platforms and the EU Digital Single Market: A response to the call for evidence by the House of Lord's Internal Market Sub-Committee," A co-production of the friends of e-Conomics. Beneficiary: House of Lords of the United Kingdom of Great Britain and Northern Ireland. Intellectual guidance: Lapo Filistrucchi., 23 November 2015. [Online]. Available: [http://ec.europa.eu/information\\_society/newsroom/image/document/2016-7/nikolai\\_van\\_gorp\\_-\\_response\\_e-economics\\_to\\_the\\_uk\\_house\\_of\\_lords\\_call\\_for\\_evidence\\_14020.pdf](http://ec.europa.eu/information_society/newsroom/image/document/2016-7/nikolai_van_gorp_-_response_e-economics_to_the_uk_house_of_lords_call_for_evidence_14020.pdf)

M. Bourreau, A. De Streel & I. Graef, *Big data and competition policy: Market power, personalised pricing and advertising*, CERRE project report February 2017, p. 30-31, available at <https://cerre.eu/publications/big-data-and-competition-policy>

J. Crémer, Y.-A. de Montjoye, H. Schweitzer (2019): *Competition Policy for the Digital Era: Final Report, Luxembourg 2019*, Publications Office of the European Union.

Data Intelligence Hub," [Online]. Available: <https://dih.telekom.net/en/>

Data Pitch - Innovation Programme, "About Data Pitch," [Online]. Available: <https://datapitch.eu/about-us/start-up/>

Datenethikkommission (2019): Gutachten der Datenethikkommission. [https://www.bmi.bund.de/DE/themen/it-und-digitalpolitik/datenethikkommission/arbeitsergebnisse-der-dek/arbeitsergebnisse-der-dek-node.html;jsessionid=7A2F84BF3737A4847ADC73A787ECFF57.1\\_cid373](https://www.bmi.bund.de/DE/themen/it-und-digitalpolitik/datenethikkommission/arbeitsergebnisse-der-dek/arbeitsergebnisse-der-dek-node.html;jsessionid=7A2F84BF3737A4847ADC73A787ECFF57.1_cid373)

"Dawex," [Online]. Available: <https://www.dawex.com/en/>

Deloitte , "Open growth: Stimulating demand for open data in the UK," 2012. [Online]. Available: <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/open-growth.pdf>

Drex1, J.(2017): Designing Competitive Markets for Industrial Data – Between Propertisation and Access, 8 (2017) JIPITEC 257-292.

Duch-Brown, N., Bertin, M., Mueller-Langer, F. (2017): The economics of ownership, access and trade in digital data. JRC Digital Economy Working Paper 2017-01.

European Commission (2017), "Staff Working Document on the free flow of data and emerging issues of the European data economy. Accompanying the document: Communication - Building a European data economy {COM(2017) 9 final}," Brussels, 10.1.2017; SWD(2017) 2 final , 10

January 2017. [Online]. Available: <https://ec.europa.eu/digital-single-market/en/news/staff-working-document-free-flow-data-and-emerging-issues-european-data-economy>

European Commission (2018): Towards a common European Data Space, <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2018:0232:FIN>.

European Commission (2018): Study on data sharing between companies in Europe, <https://op.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en>

European Commission (EC), "Statement - Digital Single Market: Commission welcomes European Parliament's vote on new rules to improve fairness and transparency of online platforms," Strasbourg, 17 April 2019. [Online]. Available: [http://europa.eu/rapid/press-release\\_STATEMENT-19-2160\\_en.htm](http://europa.eu/rapid/press-release_STATEMENT-19-2160_en.htm)

"European Data Incubator (EDI)," [Online]. Available: <https://edincubator.eu/>

European Data Market study measuring the size and trends of the EU data economy (2017) <https://ec.europa.eu/digital-single-market/en/news/final-results-european-data-market-study-measuring-size-and-trends-eu-data-economy>

European Political Strategy Centre (EPSC) (2017), Entering the Data Economy. EU Policies for a Thriving Data Ecosystem. [https://ec.europa.eu/epsc/publications/strategic-notes/enter-data-economy\\_en](https://ec.europa.eu/epsc/publications/strategic-notes/enter-data-economy_en).

Furman, Jason /D. Coyle/A. Fletcher/ P. Marsden/ D. McAuley (2019): Unlocking digital competition. Report of the Digital Competition Expert Panel, available at: <https://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel>.

Goldfarb, Avi, and Catherine Tucker (2019): "Digital Economics", Journal of Economic Literature, 57 (1): 3-43

"Google Books Ngram Viewer," [Online]. Available: <http://storage.googleapis.com/books/ngrams/books/datasetsv2.html>

"Google Finance," [Online]. Available: <https://www.google.com/finance>

"Google Trends," [Online]. Available: <https://trends.google.com/trends/explore>

P. Hausemer , L. Rabuel and H. Graux , "Study on data in platform-to-business relations - Final report," Valdani Vicari & Associati: Economics & Policy; Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs, European Commission, November 2017 (VVA (2017) [Study on data in platform-to-business relations](#))

HM Treasury (2018): The economic value of data: discussion paper. London. <https://www.gov.uk/government/publications/the-economic-value-of-data-discussion-paper>

House of Lords , "Online Platforms and the Digital Single Market," House of Lords, Select Committee on European Union, 10th Report of Session 2015-16, 20 April 2016. [Online]. Available: <https://publications.parliament.uk/pa/ld201516/ldselect/ldcom/129/129.pdf>

Hugenholtz, Bernt (2017): Data property in the system of intellectual property law, paper presented at conference “Trading Data in the Digital Economy: Legal Concepts and Tools”, Muenster Colloquium on EU Law and the Digital Economy, University of Muenster, 4-5 May 2017.

Kerber, Wolfgang (2017): Rights on Data: The EU Communication ‘Building a European Data Economy’ From an Economic Perspective.

[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3033002](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3033002)

P. Koutris, P. Upadhyaya, M. Balazinska, B. Howe and D. Suciu, "Query-Based Data Pricing," *Journal of the ACM (JACM)*, vol. 62, no. 5, 2015.

Lear (2019): Ex-post Assessment of Merger Control Decisions in Digital Markets,

[https://www.learlab.com/wp-content/uploads/2019/06/CMA\\_past\\_digital\\_mergers\\_GOV.UK\\_version-1.pdf](https://www.learlab.com/wp-content/uploads/2019/06/CMA_past_digital_mergers_GOV.UK_version-1.pdf)

P. Leonard, "The good oil on valuing "the new oil", " *Computer and Telecommunications Law Review*, vol. 24, no. 7, pp. 167-170, 2018

B. Marr , "Big Data: 33 Brilliant And Free Data Sources Anyone Can Use," *Forbes* (Online), 12 February 2016. [Online]. Available: <https://www.forbes.com/sites/bernardmarr/2016/02/12/big-data-35-brilliant-and-free-data-sources-for-2016/>

"Million Song Dataset," [Online]. Available: <https://aws.amazon.com/datasets/million-song-dataset/>

Munther Dahleh, interviewed by Elizabeth Heichler, "Why the Data Marketplaces of the Future Will Sell Insights, Not Data," *MIT Sloan Management Review*, 6 July 2018. [Online]. Available: <https://sloanreview.mit.edu/article/why-the-data-marketplaces-of-the-future-will-sell-insights-not-data/>

L. Myler , "Data Sharing Can Be A Catalyst For B2B Innovation," *Forbes* (Online), 11 September 2017. [Online]. Available: <https://www.forbes.com/sites/larrymyler/2017/09/11/data-sharing-can-be-a-catalyst-for-b2b-innovation/>

P. Nooren, N. van Gorp, N. van Eijk and R. Ó Fathaigh , "Should We Regulate Digital Platforms? A New Framework for Evaluating Policy Options," *Policy & Internet*, vol. 10, no. 3, pp. 264-301, 2018

OECD (2019): Enhancing Access to and Sharing of Data : Reconciling Risks and Benefits for Data Re-use across Societies. [https://www.oecd-ilibrary.org/sites/276aaca8-en/1/1/1/index.html?itemId=/content/publication/276aaca8-en&mimeType=text/html&\\_csp\\_=a1e9fa54d39998ecc1d83f19b8b0fc34&itemIGO=oecd&itemContentType=book](https://www.oecd-ilibrary.org/sites/276aaca8-en/1/1/1/index.html?itemId=/content/publication/276aaca8-en&mimeType=text/html&_csp_=a1e9fa54d39998ecc1d83f19b8b0fc34&itemIGO=oecd&itemContentType=book)

OECD (2015): Data-Driven Innovation: Big Data for Growth and Well-Being. OECD.

Open Data Institute (ODI), "Open enterprise: how three big businesses create value with open innovation ," London, UK, 2016. [Online]. Available: <http://theodi.org/open-enterprise-big-business>

H. Richter and P. R. Slowinski, "The data sharing economy: on the emergence of new intermediaries," *International Review of Intellectual Property and Competition Law*, vol. 50, no. 1, pp. 4-29, 2019

Rodríguez de las Heras Ballell, Teresa (2017): "The Legal Anatomy of Electronic Platforms: A Prior Study to Assess the Need of a Law of Platforms in the EU", 3 The Italian Law Journal, num. 1/3, 2017, pp. 149-176

Royal Academy of Engineering, "Towards trusted data sharing: guidance and case studies - Data sharing: implications for policy and practice," December 2018. [Online]. Available: <https://www.raeng.org.uk/publications/reports/data-sharing>

F. Stahl, F. Schomm and G. Vossen, "A classification framework for data marketplaces," *Vietnam Journal of Computer Science*, vol. 3, no. 3, p. 137–143, 2016

E. Scaria, A. Berghmans, M. Pont, C. Arnaut and S. Leconte, "Study on data sharing between companies in Europe," A study prepared for the European Commission Directorate-General for Communications Networks, Content and Technology by everis Benelux , 24 April 2018. [Online]. Available: <https://publications.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en>

G. Smith , H. A. Ofe and J. Sandberg , "Digital Service Innovation from Open Data: Exploring the Value Proposition of an Open Data Marketplace," in *49th Hawaii International Conference on System Sciences (HICSS)*, Kauai, 2016

K. Stathoulopoulos, "The rise of the platform economy: Using network science to explain the where, why and how.," Nesta, 27 June 2018. [Online]. Available: <https://www.nesta.org.uk/blog/rise-platform-economy/>

Transport Systems Catapult, "The Case for Government Involvement to Incentivise Data Sharing in the UK Intelligent Mobility Sector: Briefing Paper," March 2017. [Online]. Available: <https://s3-eu-west-1.amazonaws.com/media.ts.catapult/wp-content/uploads/2017/04/12092544/15460-TSC-Q1-Report-Documents-Suite-single-pages.pdf>

Zech, Herbert (2016): "A Legal Framework for a Data Economy in the European Digital Single Market: Rights to Use Data", *Journal of Intellectual Property Law & Practice*, Vol. 11, 460-470.