



# Final Report

Expert Group for the Observatory on the Online Platform Economy

# Work stream on Data



## Contents

<b>Executive summary</b> .....	3
<b>1. Introduction</b> .....	5
<b>2. Data: conceptualisation and classification</b> .....	7
<b>3. Data-related practices in the Platform Economy</b> .....	14
Data acquisition .....	15
Data sharing .....	19
Incentives for data sharing .....	19
Modalities of data sharing.....	21
Platforms' data sharing with their business users .....	28
<b>4. Main policy issues</b> .....	34
<i>Value of data: a premise</i> .....	34
<i>Data as a source of market power</i> .....	35
<i>Data sharing</i> .....	37
Data governance.....	39
<b>5. Conclusions</b> .....	42

# Report on Data

Author: Teresa Rodríguez de las Heras Ballell

Co-authors: Jeanette Hofmann, Inge Graef,  
Sophie Stalla-Bourdillon, Doh-Shin Jeon,  
Annabelle Gawer and Agata Majchrowska\*



## 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 and strategies by which these digital intermediaries provide these efficiencies universally revolve around the use of technology that intensively and extensively builds on data. The way data is generated and shared becomes a critical issue in a context where online services are increasingly diversified. Such data is the subject of this report. Data generated through or in relation to online platforms fosters innovation. Data plays an increasingly important role in business intelligence, product development, and process optimization.

---

\* Some sections of the report draw significantly on the following:

PPMI (2020) Analytical Paper 1: Platform data access and secondary data sources, *Support study to the Observatory for the Online Platform Economy*, available at [https://platformobservatory.eu/app/uploads/2020/09/Analytical-paper-1-Platform-data-access-and-secondary-data-sources\\_final.pdf](https://platformobservatory.eu/app/uploads/2020/09/Analytical-paper-1-Platform-data-access-and-secondary-data-sources_final.pdf)

PPMI (2020) Analytical Paper 5: Business user and third-party access to online platform data, *Support study to the Observatory for the Online Platform Economy*, available at [https://platformobservatory.eu/app/uploads/2020/09/Analytical-Paper-5-Business-user-and-third-party-access-to-data\\_final.pdf](https://platformobservatory.eu/app/uploads/2020/09/Analytical-Paper-5-Business-user-and-third-party-access-to-data_final.pdf)

<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.

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 provides 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 report is organized as follows:

Firstly, it gives an overview of the concept of data, describes how it is generated and suggests possible classifications.

Secondly, it illustrates the diversity of data sharing practices, which corresponds to the variety of business models in the platform economy. An overview of different data categories maps platforms’ data practices, notably as regards their data collection and the access they allow to the data they hold as well as limits to such data access. The report examines different approaches to data usage, access and data sharing. The report looks into various incentives and models for data sharing, assessing whether there is room to promote it 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 innovative capacities of platforms rely, in particular, on the richness of the data they collect, which allows them to gain deep insights on how the markets function, prospects and trends. 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.

Thirdly, it investigates a number of policy issues. 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 issues arise in relation to how far the access to and processing, on an exclusivity basis, 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 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. Notably, the report calls for the work on data to be continued by the Observatory of the online platform economy, beyond the mandate of the present expert group.

# 1. Introduction

By using digital technology, platforms have appeared as a new and improved mechanism to address the fundamental problem of economic organisation: how to coordinate supply and demand when information is imperfect to reach the highest possible efficiency. In the context of sophisticated economies with increasingly varied types of goods and services on offer, and a wider dispersion of preferences from the consumption side, finding valuable interactions becomes a complex task. Platforms have proposed a solution based on (matching) algorithms and data. In order to provide accurate matches, algorithms require large amounts of data, making data collection and access a key feature of the platform economy.

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, tend to take too limited views of the meaning of the term “data”, or not go deep enough in describing it, disregarding the numerous facets that data can have, depending on how and by whom it is generated and used.

To attempt to enrich the understanding of data in the platform economy, 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 also to bring focus 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. This report should be viewed as part of ongoing research of the Observatory on Data in the online platform ecosystem.

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 efficiency and competition, but also fairness, privacy and ownership. Against this background, the European Commission has just presented its data strategy for the next five years<sup>3</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. Besides, as part of the European Digital Strategy, the proposal of the European Commission of a new Digital Services Act package is aimed to address the profound change of the digital services market in a platform economy. The expansion and relevance of online platforms in the provision and intermediation of digital services have severely challenged the legal framework for digital services, unchanged since its adoption in 2000. The recently adopted Digital Services Act (DSA) will upgrade the ecommerce framework for all internet services in the EU to ensure that digital services face no borders inside the Union,

---

<sup>3</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)

while defining clearer responsibilities and accountability for online platforms such as social media and marketplaces for preventing illegal content and goods online.

The other legislative proposal recently put forward by the European Commission, the Digital Markets Act (DMA), will ensure that markets characterised by large platforms acting as “gatekeepers” for businesses to reach consumers, remain fair and contestable for innovators, businesses, and new market entrants. It will establish rules to ban a number of unfair and anticompetitive practices used by such gatekeeper platforms exploiting their unprecedented economic power to the detriment of (smaller) businesses and consumers.

These recent policy initiatives in the EU are aiming to touch upon some data-related aspects, but they will not exhaustively address all issues in the data economy. A step-by-step approach is required in this regard, given the multiple implications, some of which unforeseen or undesirable, that measures in this area can have. To ensure the soundness of such an approach, one of the indispensable steps is to recognize the complexity of the matter, and attempt to understand it to the greatest possible extent.

For that reason, the analysis in this report aims to move away from general – and, implicitly, imprecise – statements on the need to use, access, and 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 in the online platform economy. The survey reveals that the variety of data by platforms and the constellation of data practices implemented by platforms require different data governance solutions and a more granular policy approach to balance interests at stake and achieve policy goals.

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? Which categories of data are primarily collected, observed, and inferred by platforms? Does the value of data depend on sectors? Which data have higher value in the platform economy? Which incentives should platform have to collect and process data if they have to share it with competitors or third parties? The answers to those questions are important to guide policy makers.

Data sharing can be internal or external<sup>4</sup>. Data can be shared within the platform, between business users and the platform operator, which collects, observe and infers data from users’ activity. Platforms can also share data with third parties, either competitors or data seekers, or even government authorities. Incentives for the platform to share data internally or externally differ, as well as the legal, market, and policy consequences. Data sharing can be also voluntary trade in data (remunerated), voluntary exchange without remuneration, or mandatory sharing. Sharing can be direct (a transfer of data between two parties, or indirect (trading or sharing a

---

<sup>4</sup> JRC (2020), Digital Economy Working Paper 2020-5, *B2B Data sharing: An economic and legal analysis*.



data-based service, insights derived from data but not the underlying primary data). Policy makers have to take these different angles of data sharing into consideration in devising data governance solutions.

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 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 usage and sharing are treated in recent and ongoing work: for instance, the 2019 Report by Special Advisors to the Competition Commissioner Margrethe Vestager<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. Data: conceptualisation and classification

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, for instance. Consequently, technical, economic and legal definitions of data need further context and

---

<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 (2019), *Competition policy for the Digital Era*, Publications office of the European Union.

[https://ec.europa.eu/competition/information/digitisation\\_2018/report\\_en.html](https://ec.europa.eu/competition/information/digitisation_2018/report_en.html)

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>

### *Different conceptualisations of data*

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 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 expressed, “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 approach focuses on the properties or affordance of data, understood as lending themselves to specific forms of analysis<sup>11</sup>. They acquire their value as relational entries of data sets, which, in turn, may become components of other, larger data sets. Although not all data is valuable, their main characteristic is that they can be “used, reused, copied, moved, and processed cheaply, without degradation, at very fast speeds, and globally”<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>.

From an **economic** perspective, the key characteristic of data is non-rivalry: 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>. As non-

---

<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.

<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).



rivalry implies that the same data can be used by a multitude of actors for a variety of purposes without causing functional loss to the original data collector or subsequent data user, significant potential gains can be attributed to data-sharing and data-re-use policies. This implies that social welfare will be maximised when access to the non-rival good is also maximised, as each extra private benefit comes at no extra cost<sup>17</sup>. The same feature of non-rivalry, in connection with the economies of scope, does also reveal another source of economic efficiency gains and social welfare: data aggregation. Merging two complementary datasets can generate more insights and economic value compared to keeping them separate.

Different from physical goods, data are not excludable by nature: they can be copied and disseminated. Although the law can assign exclusive rights to data originators and/or collectors, they tend to use different strategies to guarantee an exclusive control and access to the data. The non-rivalry together with the excludability or non-excludability of data<sup>18</sup> are characteristics that data governance models must consider and internalize. A need of balance or trade-off between the social welfare of wider data access and the benefits and efficiency of data concentration should guide policy decisions.

The **legal** approach to data or data-related practices focuses on rights relating to use, curation, control, stewardship or exclusive access. Despite the lively debate and the existence of certain ownership-like rights in sectoral legislations on intangible assets, the legal approach does not pivot on an explicit recognition of data ownership rights.

Specific properties of data affect the legal framework of "data governance"<sup>19</sup>. For the law to regulate data, it has to specify what it intends to protect<sup>20</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>21</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>22</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>23</sup>, the debate has evolved in the meantime.

---

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

<sup>18</sup> Being non-rival, data can be either non-excludable (and then a public good) or excludable (hence a club good).

<sup>19</sup> See OECD (2015: 186)

<sup>20</sup> Drexl (2017: 263)

<sup>21</sup> OECD (2015: 195)

<sup>22</sup> OECD 2015: 195-7, with regard to patient data in the health sector; see also Drexl (2017: 260)

<sup>23</sup> Croll A. (2011) quoted in OECD (2015: 197)

To conclude, the singular properties of data and the purpose the data will be used for allow for a variety of data governance models, which differ according to the distribution of control and access and their overall social welfare effects.<sup>24</sup>

### *Classification of data categories in platforms*

The wide variety of data-related practices in the platform economy reflects the great diversity of data, the variation in the value of data in different sectors, and the heterogeneity of data collection, data usage, and data processing policies. Any policy approach to regulate data access or data sharing needs to consider this variety and heterogeneity of data and data-related practices, while should pursue specific purposes<sup>25</sup>. Therefore, the analysis of data-related practices in the platform economy requires some prior classification of data categories, 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>26</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 or stock prices. The borderline between the two categories is becoming increasingly blurred, in particular in an IoT context<sup>27</sup> while re-identification technologies are becoming more and more sophisticated<sup>28</sup>. Mixed data sets, including both personal and non-personal data bring additional complexities; the European Commission issued a set of

---

<sup>24</sup> OECD (2015) Data-Driven Innovation: Big Data for Growth and Well-Being, (195-7)

<sup>25</sup> As stated in the better regulation guidelines of the EC, for instance.

<sup>26</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>27</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>28</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”).

guidelines aiming at adding further clarity as to the legal requirements concerning such data sets<sup>29</sup>. Ongoing research in areas such as differential privacy aim at reconciling data sharing and privacy concerns<sup>30</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 anonymisation.<sup>31</sup> This consideration creates tensions with the traditional static approach to the scope of the data protection framework<sup>32</sup>.

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.

A classification commonly used for personal data describes three categories according to its provenance: **volunteered data**, **observed data** and **inferred data**<sup>33 34</sup>. “Volunteered data”<sup>35</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, clicks on products. Finally, “inferred data” is information that a platform actively creates by aggregating, processing and analysing volunteered or observed data (by use of data analytics such as clustering, filtering or prediction), e.g. customer risk profiles or analysed user sentiment. Despite the utility of this classification, it must be acknowledged that in certain cases the distinction among the different categories of data – volunteered, observed, and inferred – blurs (e.g. product review scores). Such added complexity in the application of the taxonomy should be taken into consideration in the analysis of data-related practices.

---

<sup>29</sup> <https://ec.europa.eu/digital-single-market/en/news/practical-guidance-businesses-how-process-mixed-datasets>

<sup>30</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.

<sup>31</sup> Hence the importance of operating within a controlled environment.

<sup>32</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>33</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>34</sup> *Data sharing for digital markets contestability: towards a governance framework*, CERRE report September 2020, p 16-17, available at <https://cerre.eu/publications/data-sharing-digital-markets-competition-governance/>

<sup>35</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”.

The collection of each category of data entails different costs for the platform, with notable economies of scale. Whereas the collection of volunteered data is mainly based on the provision by the data users, observed data and inferred data imply collecting and processing costs. Consequently, data sharing and re-use costs as well as incentives for data sharing differ depending upon the category of data at stake. From a purely supply-side cost perspective, the higher the costs and the value of data for the platforms are, the lower the incentives to share data would be. Nevertheless, taking into account the demand side value, platforms would have incentives to share high-value data precisely due to the potential high price in data trade (e.g. consumer profiling and targeting data on advertising platforms). Thus, in deciding the data-sharing policy, platforms assess the expected benefits and costs.<sup>36</sup>

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>37</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).
- **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").<sup>38</sup>

---

<sup>36</sup> JRC (2020), Digital Economy Working Paper 2020-5, *B2B Data sharing: An economic and legal analysis*.

<sup>37</sup> VVA (2017) [Study on data in platform-to-business relations](#), p. 27.

<sup>38</sup> From the perspective of the platform business users, the following additional distinction is important: i) Data on the business's own listings, such as sales, customers, transactions and business performance. It is used by businesses to manage the accounting, stocks, sales and delivery, forecast future distribution needs and so on; ii) Data on the competitor or overall listings, customers, transactions and performance of businesses on the platform in a specific market. This kind of data (e.g., search keywords, search volumes, consumer behaviour, etc.) is

The above-listed categories of data, whose relevance is confirmed by market research,<sup>39</sup> can be complemented by another classification criterion based on the source of data. Data on individual transactions, business performance and market trends are collected by platforms from all business users operating in the platform. Data related to its own business are relevant for the business users for planning and managing all business activities. But overall data related to business performance, consumer behaviour, or market trends of all the other business users of the platforms provide critical additional information and add significant value for business decision as they provide an overview of processes, activities, trends, and competition in the platform. Such market- or platform-wide data, that can easily include commercially valuable information likely to raise trade secrets considerations, are less frequently shared by platforms to users and to a very limited extent.

As these lists indicate, 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>40</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>41</sup> The value of a sectoral approach has been flagged also by the stakeholder feedback received in response to the initial version of the report, of July 2020.

In addition, depending on the 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>42</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

---

collected by the platforms, that therefore gain the full overview of the activities taking place on it. For the business users, studying competition is one of the key steps to succeed on online platforms. See Gerber, S. (2018), "Want to sell on Amazon? 15 strategies for success", *The Next Web (TNV)*. Available at <https://thenextweb.com/contributors/2018/01/19/want-sell-amazon-15-strategies-success/>

<sup>39</sup> PPMI (2020), *Platform data access and second data sources*, Analytical paper 1, p. 18.

<sup>40</sup> Crémer, Montjoye and Schweitzer (2019), *Competition policy for the digital era*, p. 25 et seq

<sup>41</sup> VVA (2017) [Study on data in platform-to-business relations](#), p. 35.

<sup>42</sup> Schweitzer et al. (2018), *Modernisierung der Missbrauchsaufsicht für marktmächtige Unternehmen*, p. 60 et seq.

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>43</sup> and foreclosing competitors in these markets.

Further criteria to be taken into account for classifying data 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. Incentives to share and capacity to benefit from data vary for business users depending on these additional characteristics. Disaggregated and raw data are not highly valuable if business users lack capacities to process unprocessed data. Analytics and data-based insight require capacities and competence. Therefore, the value for users is higher, but the incentives to share by platform are lower, as resources and costs have been invested and competitive advantages can be gained from private use.

Finally, broad considerations about the source of data should also be accounted for<sup>44</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.

### 3. Data-related practices in the Platform Economy

This report lies on the premise that platforms’ innovative capacities rely, among others, on the richness of the data they collect. This is a unique feature of these digital services. For instance, e-commerce platforms are able to collect abundant data on their business users as well as on customers’ previous purchases, reviews, preferences, and search histories. These data allow platforms to continuously adapt and adjust their service<sup>45</sup>. The 2019 expert group report<sup>46</sup> stressed the online marketplaces’ unique position with regard to data: They “can observe in detail how the market functions”; collect and analyse not only customer preferences but also the complementarity of products, how prices adjust and other relevant variables to monitor how the market is evolving.

---

<sup>43</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 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>44</sup> Data provenance is an active area of research, which includes efforts for a standardised [provenance](#) framework

<sup>45</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>46</sup> Crémer, Montjoye and Schweitzer (2019), Competition policy for the digital era, p. 68.



This section intends to give a broad overview of data-related practices in the platforms' environment.

This mapping and analysis exercise should be read as a non-exhaustive snapshot of the main types of behaviors, trying to illustrate a much vast – 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>47</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 by other participants in the broader ecosystem. 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 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 as it is based on data generated or accumulated by individual competitors 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 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>48</sup>.

The cost (investment) needed to collect or produce the data is another relevant consideration. Here the **distinction between volunteered, observed and inferred data**, is relevant. Firstly, online platforms create the environment and maintain the service that leads to the emergence of “digital data traces” from their users, their transactions and their online behaviour. While certain data classified as “volunteered” have to be provided as a prerequisite to use the services,

---

<sup>47</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>48</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>.

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 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.

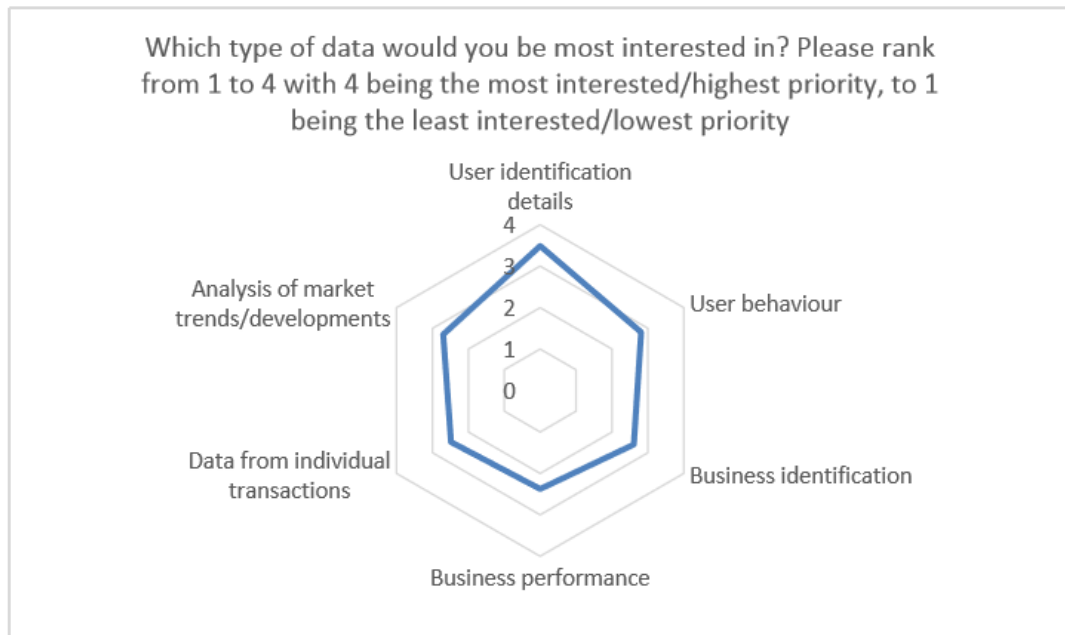
Conversely, inferred data is information that platforms create themselves by analysing volunteered and observed data. Therefore, this type of data probably requires the most significant investment from platforms. While the detection of regularities and patterns in quantitative data may already require 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 by means of 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 platforms' relative competitive advantages. 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.

**Business users of online platforms** generally – but not always – cannot match the data-driven innovation capacity of online platforms. Empirical analysis<sup>49</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 business users seek greater access to customer data out of an interest in direct contact of consumers, disintermediating the transactions.

---

<sup>49</sup><https://publications.europa.eu/en/publication-detail/-/publication/4af6cec1-48fb-11e8-be1d-01aa75ed71a1/language-en>



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

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>50</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).

The limited empirical evidence available shows that four-fifths (81%) of the respondents to a platforms' business users survey indicated that they collect some data themselves. The most prevalent data type refers to identification details of own customers (57% of respondents collect this data), followed by business performance data (55%) and analysis of market trends/developments (55%). Further, a third of the surveyed companies (33%) reported that they use third-party sources (data brokers). Most of the interviewees - especially in the e-commerce sector - reported that they use the services of third-party data and analytics providers<sup>51</sup>.

<sup>50</sup> See chapter on categories of data, *supra*.

<sup>51</sup> PPMI (2020), *Platform data access and second data sources*, Analytical paper 1.

Companies specialising in data and insights fill important data gaps, especially with regard to data on competition and actionable business insights.<sup>52</sup> Some data brokers specialise to cover specific sectors, such as applications (App Annie, App figures, Apptopia, Mobile Action, Sensor Tower), e-commerce (Algopix, Jungle Scout, Sellics, Teikametrics, Terapeak) or hospitality (AirDNA, Beyond Pricing, Uplisting, Wheelhouse, Skift). Other data brokers, such as Similar Web and Zirra provide data on multiple sectors.

Data brokers use highly advanced technical methods to extract data, or they buy data from online and offline sources. A lot of data is scraped from the platforms.<sup>53</sup> While some platforms do not permit scraping for commercial purposes that does not always prevent data companies from collecting data this way.

Another key source is crowdsourcing business user account data. Some third-party data providers ask business users to share their marketplace information, and then link the data of thousands of users to draw market insights.<sup>54</sup> When merged with the data gathered by scraping the platform's front-end (e.g., Best Seller rank on Amazon), this can yield quite precise estimations and extrapolations.

The key value proposition of the data brokers lies in their ability to bring together a combination of sources as well as superior technical and analytical capacities, innovative tools and approaches. Data brokers allow their users to learn about their competitors, get a detailed market overview, and obtain actionable insights.<sup>55</sup> According to analysts<sup>56</sup>, this is the kind of information that is most in demand by the business users that platforms do not provide to a sufficient extent. Further, business users themselves do not need to invest into any analytics or IT, but rather buy products tailored to their needs.

The data companies' market is very dynamic and fast-paced. A number of issues exist, illustrating the key challenges and limitations of data brokers. Firstly, the data companies

---

<sup>52</sup> According to PwC estimations, data companies earned \$21 billion in 2018. Gröne, F., Péladéau, P., & Samad, R. A. (2019). Tomorrow's data heroes. Strategy+business. Retrieved from <https://www.strategy-business.com/article/Tomorrows-Data-Heroes?gko=5f270>

<sup>53</sup> Matsakis, L. (2019). Amazon cracks down on third-party apps over privacy violations. Wired. Retrieved from <https://www.wired.com/story/amazon-marketplace-apps-privacy/>

<sup>54</sup> For example, Jungle Scout collects data from a large number of sellers (over 225,000) who have opted in to share their sales information. Rohler, M. (2019). How are Estimated Sales and Revenue Calculated? What are AccuSales? Jungle Scout Help Center. Retrieved from <https://support.junglescout.com/hc/en-us/articles/360008616814-How-are-Estimated-Sales-and-Revenue-Calculated-What-are-AccuSales-#:~:targetText=We%20gather%20it%20from%20a,their%20sales%20info%20with%20us.&targetText=This%20means%20that%20as%20we,analyze%20and%20test%20it%20daily>. Similarly, if AirDNA users wish to receive performance analytics, they will be asked to upload their Airbnb host IDs. After doing this, they can see their performance trends, comparative and financial analysis on all vacation rental listings.

<sup>55</sup> In the accommodation/ hospitality sector, companies such as AirDNA provide insights based on data that the OTAs do not share. For example, in late 2015 Airbnb stopped providing the overall real-time reservation data. AirDNA, in turn, uses an algorithm based on 16 indicators picked up in historical data to determine the reservation status for each listing. They argue that their algorithm has an error margin of only 5%.

<sup>56</sup> PPMI (2020), *Platform data access and second data sources*, Analytical paper 1.

remain highly dependent on data sharing policies of platforms.<sup>57</sup> Secondly, the data needs of platform business users are often very specific and concern platforms that they use. Such data cannot be easily scraped or estimated by the third-party data providers<sup>58</sup>. It includes information on real-time of activities on the platform, which would allow to effectively address the customer; transaction-related data about the customers, sales activities and listings of specific business user. Thirdly, the huge amounts of data that data brokers collect, store, possibly re-personalise and disseminate are of interest from the regulatory perspective, first and foremost due to privacy concerns. Most individuals or companies are unaware of what information data brokers collect on them or even that they collect information at all.<sup>59</sup> Due to this asymmetry, the data broker industry has been often characterised as opaque, non-transparent, arbitrary, biased, unfair and unaccountable<sup>60</sup>.

As a final point it is also observed that some business users are exploring innovative approaches that would allow them to joint forces and be less dependent on big platform companies. One example includes cooperative marketplaces, such as Fairmondo.de, which belongs to its business users and employees. Through a cooperative structure, the users can share the platform as a resource for mutual benefit and decide on the rules for data sharing and access.

## Data sharing

### Incentives for data sharing

The incentives around data sharing are complex and can be related to the business strategies of the actors, the objective pursued by, and the potential value. 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.

Data sharing practices are at the core of the data(-driven) economy and are 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

---

<sup>57</sup> For example, Amazon until recently provided exact and broad match search volume and product relevance data via one of its APIs. It was feeding several third-party software providers such as Viral Launch and Helium10 until late 2018, when the platform removed these metrics from the API. Another platform, Allegro has recently made significant investment to develop new data products (Allegro Statistics) that are now provided to its sellers; this is endangering the business model of third-party analytics providers.

<sup>58</sup> Graef, I. (2016). EU Competition Law, Data Protection and Online Platforms: Data as Essential Facility. Kluwer Law International BV.

<sup>59</sup> Christl, W. (2017). How companies use personal data against people. Working paper by Cracked Labs, Institute for Critical Digital Culture.

<sup>60</sup> Christl, W. (2017). How companies use personal data against people. Working paper by Cracked Labs, Institute for Critical Digital Culture.

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>61</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.

When taking decisions to share or not to share data, online platforms must reconcile several competing and potentially conflicting imperatives. On the one hand, the success of the business users is important because it generates revenues for the platform. In this sense, online platforms have a strong incentive to provide access to data that could help businesses to understand their customers and to improve their product. On the other hand, online platforms must maintain trust of their clients (business users and customers of the business users), which means that they should avoid sharing data that these clients are unwilling to share, for example, personal information, sensitive business information.

Online platforms have also designed their terms and conditions to comply with the applicable regulatory frameworks, including P2B regulation, personal data protection, competition law, regulation forbidding trade in illegal and counterfeit products, and others.

Several groups of players operate within the data ecosystem surrounding each online platform. These include other platforms, large and small businesses, customers of the business users, data brokers or companies providing online optimisation tools, regulatory and other public authorities. Sometimes these groups have diverging interests and competing demands concerning data access. As mentioned earlier, the platforms see personal data protection as part of their value proposition, however this claim is not always accepted by some businesses who argue that platforms use data protection as an excuse for not sharing important data. If platforms

---

<sup>61</sup> J. Crémer, Y.-A. de Montjoye, H. Schweitzer (2019), *Competition Policy for the Digital Era: Final Report*, p. 68, 85, 96.



decide to open more raw data to business users, this could benefit large businesses at the expense of the smaller ones, because the big companies have the necessary infrastructure and know-how to take advantage of such information.

If a specific dataset is at the core of a platform's business model, it is unlikely to be shared. Due to this reason platforms will be reluctant to share datasets that could be used to undermine their role as leading intermediaries in two-sided markets. Vertically integrated platforms are not likely to share detailed market-level data, which could help the emergence of new competitors in their market. Yet these platforms also make internal decisions on what information from their marketplace/app store can or cannot be shared with the retail/app development division. Such decisions are of crucial importance to many businesses that compete with goods and services sold by the platform itself.

Finally, the lack of technical interoperability between different platforms is also a constraint impeding data sharing and data portability. Introducing interoperability is costly, because it requires the development of common standards and revision of back-end code. From the perspective of platforms, investing into interoperability does not necessarily provide a clear commercial gain. Interoperability also has its downsides because it may make the system slower and limit the development of new or innovative products.

#### Modalities of data sharing

In this section, various **models of data sharing** are described. 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 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 continuous access.* Depending on the data user needs, both models are legitimate and appropriate in different situations. Continuous, sometimes in real-time, 'data hose' established through a performance APIs is typically a precondition for uninterrupted 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>62</sup>), but less so in cases the recipient does not have the data analytics skills and the infrastructure to make use of the opportunity. One-off data sharing normally includes historical data that can be used for market research purposes.

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

---

<sup>62</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/>

Protection Regulation offers a one-off access to data only<sup>63</sup> <sup>64</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>65</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)<sup>66</sup>.

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

The Twitter social network provides access to several types of APIs facilitating access to different functionalities of the platform<sup>67</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. 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>68</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.

---

<sup>63</sup> As suggested in Crémer/De Montjoye/Schweitzer (2019), *Competition policy for the digital era*, p. 81-82.

<sup>64</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>65</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>66</sup> The most eminent example of self-regulation that relates to innovation and competition is the governance of standard essential patents (SEPs) and their licensing via FRAND commitments in the standard setting context. The FRAND obliges patent holders to grant licenses to interested parties under Fair, Reasonable, and Non-Discriminatory terms (what this actually means in particular situations are often defined in antitrust court cases). Originally, the rationale to introduce a requirement to grant FRAND-compliant licenses was based purely on competition law and prevention of abuse of market power. Currently there is also a contractual basis for a FRAND license, since most standard-setting organisations require patent holders to grant licenses on FRAND terms in order for a patented technology to be included in a standard.

<sup>67</sup> <https://developer.twitter.com/en/products/products-overview>

<sup>68</sup> See e.g. <https://www.programmableweb.com/news/how-datasift-survived-twitters-merciless-business-behavior-api-economy/native-case-study/2018/12/11>

In the 2017 Communication "Building a European Data Economy", the Commission explored the idea of applying a FRAND regime in relation to access to data, whereby access to machine-generated data would be granted against remuneration<sup>69</sup>. The Communication notes that a framework based on FRAND terms could be developed for data holders to provide access to the data they hold against remuneration after anonymisation. Relevant legitimate interests, as well as the need to protect trade secrets, would need to be considered.

The constraints of platform data sharing in some respects are similar to those that FRAND aims to solve. First, concentration of data in the hands of several large platforms can potentially lead to the abuse of market power. Data can be essential for conducting business in certain markets and function as gatekeepers. Any service or product requiring these data is dependent on the goodwill of the data holder. Second, opening more data remains a sensitive question for platforms. FRAND, in turn, is seen in literature as a measure to increase transparency and trust among the participants and therefore incentivise more data openness<sup>70</sup>.

*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>71</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>72</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>73</sup>

It could be argued that data held by dominant online platforms is an “essential facility” that rivals (other platforms or businesses) need access to in order to compete in the market. It is increasingly clear that access to data, and the ability to analyse and glean insights from the information, can confer a competitive advantage. However, in the fast-moving digital markets it can be difficult to define precisely what is the relevant market, and to then prove that a specific platform is dominant in it. Competition authorities would also need to build up a portfolio of evidence to show that data held by dominant online platforms is an “essential facility” and that platforms denying access to that data are acting in breach of the competition

---

<sup>69</sup> European Commission, Towards a common European data space (Communication), COM(2018) 232 final.

<sup>70</sup> Richter, H., & Slowinski, P. R. (2019), “The data sharing economy: on the emergence of new intermediaries”, *IIC-International Review of Intellectual Property and Competition Law*, 50(1), 4-29.

<sup>71</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>72</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>73</sup> See, more in detail, J. Crémer, Y.-A. de Montjoye, H. Schweitzer (2019), *Competition Policy for the Digital Era: Final Report*, p. 83,84.

law.<sup>74</sup> Mandatory data sharing could impede the development of new business models in such a dynamic environment<sup>75</sup>, and may incentivise platforms to engage in strategic behaviour, for instance, by adding noise or large amounts of non-material and raw information of little value in the public disclosure.<sup>76</sup> All these considerations should be taken into account in assessing the implications of a general compulsory access or a specific access mandate.

*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.

**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>77</sup> that the practice of scraping publicly available data did not infringe the American Computer Fraud and Abuse Act of 1986 (CFAA).

*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>78</sup>, with an aim

---

<sup>74</sup> Graef, I. (2019). Rethinking the essential facilities doctrine for the EU digital economy. *RJT ns*, 53, 33.

<sup>75</sup> Richter, H., & Slowinski, P. R. (2019), “The data sharing economy: on the emergence of new intermediaries”, *IIC-International Review of Intellectual Property and Competition Law*, 50(1), 4-29.

<sup>76</sup> Easley, D., Huang, S., Yang, L., & Zhong, Z. (2018). The Economics of Data. Available at SSRN 3252870.

<sup>77</sup> <http://cdn.ca9.uscourts.gov/datastore/opinions/2019/09/09/17-16783.pdf>

<sup>78</sup> Antti Poikola, Kai Kuikkaniemi & Harri Honko, *MyData* (2014), *A Nordic Model for human-centered personal data management and processing*.

to empower data subjects to make use of their portability right on personal data<sup>79</sup>. Nonetheless, the success of these models are not undisputed, as it is alleged that the value of data is lower than the investment cost; consequently, they may remain marginal and hardly scale.<sup>80</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>81</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>82</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, and the community behind it continues experimentation under a Social Web Incubator Community Group<sup>83</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>84</sup>. Further discussions emerge, however, on

---

<sup>79</sup> See, for example, the P3C Consortium and the Privacy Tech White Book <https://www.privacytech.fr/livre-blanc-privacytech.pdf>

<sup>80</sup> Krämer, J., Senellart, P., and de Streel, A. (2020), *Making data portability more effective for the digital economy: Economic implications and regulatory challenges of the portability of personal data in the digital economy*. CERRE Policy Report.

<sup>81</sup> <https://datatransferproject.dev/>

<sup>82</sup> [www.w3.org/TR/social-web-protocols/](https://www.w3.org/TR/social-web-protocols/)

<sup>83</sup> <https://www.w3.org/wiki/SocialCG>

<sup>84</sup> E.g. location data, and models can also remember training data, see <https://arxiv.org/pdf/1709.07886.pdf>

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>85</sup>.

**Example: The OPAL project (Open Algorithm)**<sup>86</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 and other organisational solutions.* 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.

Data brokers and online optimisation tool providers play an important role in data markets by offering data which is not accessible directly from the platforms. They usually pool platform data from multiple sources, including publicly available data, crowdsourced business user account data, data provided by platforms through APIs and data scraped from platform websites. Platforms argue<sup>87</sup> that they do not have direct contractual relationship with the data brokers/online optimisation tool providers and thus are not responsible for quality or accuracy of the data. Nevertheless, the platforms see value in this market because it is useful for their business users. However, they may take action if, for example, they see that traffic from online optimisation tools providers start interfering with platforms' services.

---

<sup>85</sup> INSEE offers this service, see [www.casd.eu](http://www.casd.eu)

<sup>86</sup> <https://www.opalproject.org/>

<sup>87</sup> PPMI (2020), *Business user and third-party access to online platform data*, Analytical Paper 5.



A **data pool** is a centralized repository of data where different organizations can obtain, maintain, and exchange information in a standard format. **Data trusts** also function based on similar principles to data pools (see the box below)<sup>88</sup>. Companies can use data pooling to share data “in reference to a given service or generally in an industry, or within an e-ecosystem”<sup>89</sup>. In this situation, companies share a commodity (patented technology or data) to their common economic benefit. However, the exclusion of third parties or the unequal treatment of such third parties can result in an enhanced market position. Due to this additional market power stemming from the shared data, companies may be mandated to share the data with parties outside of the pool. However, the data itself in this case should be considered critical for a specific purpose<sup>90</sup>.

Moreover, in considering the case of mandated data access, the openness of the data pool is important. If the data pool is generally open for third parties to join, the rules of the pool should provide for equal non-discriminatory treatment of all participants or users. This principle can increase transparency and trust among the participants and therefore incentivize market participants to engage in sharing. If, in contrast, the pool is closed, it is a priori difficult to justify why non-participants should have a right to access the data shared within the pool.<sup>91</sup>

Another mode of data sharing is a **data sharing platform**<sup>92</sup>, which can take many forms. If the data sharing platform merely provides a **neutral marketplace** for participants to directly exchange data in a peer-to-peer model, this scenario appears similar to that of two companies sharing or exchanging data **directly**. In such cases, it seems reasonable to apply the same rules as for direct data sharing. An example could be data marketplaces (e.g. Dawex<sup>93</sup>) that provide the technical infrastructure for the exchange of data between multiple parties. From an economic perspective, their key function is to facilitate data sharing by lowering transaction costs through combining different data sources and matching users and suppliers. If, however, the platform provides a gateway to data commonly shared or exchanged between **numerous companies**, it resembles a data pool.

**Example: DAWEX<sup>94</sup>**

DAWEX<sup>95</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

<sup>88</sup> BPE Solicitors, Pinsent Masons, Reed, C., (2019), “Report on Data trusts: legal and governance considerations”. Available at <https://theodi.org/wp-content/uploads/2019/04/General-legal-report-on-data-trust.pdf>

<sup>89</sup> Lundqvist, B. (2018). Competition and data pools. *Journal of European Consumer and Market Law*, 7(4), 146-154.

<sup>90</sup> BPE Solicitors, Pinsent Masons, Reed, C., (2019), “Report on Data trusts: legal and governance considerations”. Available at <https://theodi.org/wp-content/uploads/2019/04/General-legal-report-on-data-trust.pdf>

<sup>91</sup> Richter, H., Slowinski, P. R. (2019), “The data sharing economy: on the emergence of new intermediaries”, *IIC-International Review of Intellectual Property and Competition Law*, 50(1), 4-29.

<sup>92</sup> Richter, H., & Slowinski, P. R. (2019), “The data sharing economy: on the emergence of new intermediaries”, *IIC-International Review of Intellectual Property and Competition Law*, 50(1), 4-29.

<sup>93</sup> <https://www.dawex.com/en/>

<sup>94</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>95</sup> <https://www.dawex.com/en/>

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 their business users

Platforms provide data to their business users, which is sufficient to process transactions and manage their business. The businesses receive data about their own listings, prices, sales, transactions and business performance. Platforms also provide some data about direct customers. Further, most major platforms share some data about the broader market, including overall market trends, best-selling products, customer profiles, although the type and granularity of such information differs from platform to platform. The major platforms compete for their business users and thus various metrics and dashboards are part of their value proposition. These metrics and dashboards are designed to help the business users to know their customers, monitor their own business performance, and understand the broader market trends. Table 1 provides a non-exhaustive overview of the key types of data access models.

A key dimension in discussing the usefulness of data is access to raw vs. processed data. Access to raw data is mainly of interest to the bigger business users, who have the infrastructure and specialised analytical skills to generate market insights. Yet, the majority of businesses do not have the capacity to derive value from unprocessed data. Therefore, as corroborated by the Observatory activities<sup>96</sup>, the majority of business users need processed data or data analytics that offer insights on how to better position or market their products and services, appear more frequently in search results, and communicate with clients, among others. In fact, several platforms provide data-based guidelines, tips, and advice to business users. Some examples of such advice include: optimisation of product titles based on search data; development of a new product based on predictions on what consumers will buy; adding tea kettles to hotel rooms to attract Asian tourists; change the monetisation model of an app based on user engagement trends. This is appreciated by business users.

Finally, some data may be collected by platforms and provided in both raw and processed form, but only as part of other services and for a price. For example, eBay offers the Terapeak Research product to its sellers with a Basic, Premium, Anchor or Enterprise Store account; the sellers with the Starter account may access data for a yearly or monthly subscription. Other platforms such as Amazon, Etsy and others offer advertising services that generate data, which is made available specifically to businesses subscribing to such services.

Several activities of the Observatory have shown that a significant share of business users express dissatisfaction with regard to the level of data access provided to them by online platforms. A business user survey showed that access to data possessed by online platforms is of concern to around a third of surveyed business users who reported that they cannot access at least some data that is essential to their business. Generally, although the platforms collect and analyse loads of data, only a fraction of this is provided to other players. The platforms do

---

<sup>96</sup> PPMI (2020), *Platform data access and secondary data sources*, Analytical Paper 1.

not share the raw big data on day-to-day activities, as well as detailed data on customers and competitors.

This work has identified three groups of concerns that business users express with regard to data sharing. The first is related to lack of access to personal data, such as customers' e-mail address. Some business users, especially in the hospitality and e-commerce sectors consider such data of key importance to them so that they could establish a more direct client relationship. This data is not provided for a number of reasons: business models, data protection or regulatory constraints for advertising. The second corresponds to the claim that business users need data that help them to stay competitive, innovate and develop their products and services. Partly this is related to data on, for example, search keywords, search volumes, consumer behaviour in reaction to different price signals. Platforms do share such data to a certain extent (e.g. three most important key words). This is also related to data about competitors and their products and services. The third concern is that platforms are taking advantage of data to promote their own products that are very similar to those offered by their business users. This is primarily pertinent to vertically integrated platforms with significant market power. The results of the business user survey indicate that 58% of respondents reported that the platform itself offers the same (or very similar) goods or services to those that their businesses offer on the platform. The key ways of favouring include ranking, placement of advertisement, pricing and other – all of these are enabled by the data collected by platforms.<sup>97</sup>

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 P2B 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.

---

<sup>97</sup> See PPMI (2020), *Platform data access and secondary data sources*, Analytical Paper 1. For more information, consult The Washington Post (2019). Amazon sellers say online retail giant is trying help itself not consumers. Retrieved from <https://www.washingtonpost.com/technology/2019/10/01/amazon-sellers-say-online-retail-giant-is-trying-help-itself-not-consumers/?arc404=true>; Creswell, J. (2018). How Amazon steers shoppers to its own products. The New York Times. Retrieved from <https://www.nytimes.com/2018/06/23/business/amazon-the-brand-buster.html> Hu, K. (2019). Revealed: how Amazon uses seller data to build a private label juggernaut. Yahoo! Finance. Retrieved from <https://finance.yahoo.com/news/amazon-uses-thirdparty-sellers-data-to-build-private-labels>

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.  Also referred to as “ <i>an open data approach</i> ” <sup>98</sup> , and “ <i>open data policy</i> ” <sup>99</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>100</sup> (e.g. a small payment is required to cover re-formatting costs <sup>101</sup> ).	An online platform (a) may provide open data where it has “ <i>a strong interest</i> ” in data re-usage <sup>102</sup> – e.g. for innovation purposes <sup>103</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>104</sup> . An online platform (b) can be <b>legally required</b> to make certain data openly accessible <sup>105</sup> .	E.g. Google Trends <sup>106</sup> , Google Books Ngram Viewer, Google Finance, Amazon Web Services Open Data, Million Song Dataset. <sup>107</sup>

---

<sup>98</sup> European Commission (2018): Towards a common European Data Space p. 5

<sup>99</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>100</sup> European Commission (2018): Towards a common European Data Space p. 5

<sup>101</sup> E. Scaria et al., p. 64

<sup>102</sup> European Commission (2018): Towards a common European Data Space p. 5

<sup>103</sup> E. Scaria et al., p. 64

<sup>104</sup> European Commission (2017), p. 14

<sup>105</sup> E. Scaria et al., p. 64

<sup>106</sup> European Commission (2017)

<sup>107</sup> For further examples of open datasets see B. Marr (2016)

ii.	Data monetisation model <sup>108</sup> .	An online platform makes data accessible via a “ <i>unilateral approach</i> ” – e.g. through a particular licence <sup>109</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>110</sup> .	E.g. “Telefónica [...] provides access to anonymised or aggregated insights derived from the data that the company holds [...]” <sup>111</sup> .
iii.	Data marketplace model. <sup>112</sup> Also referred to as “ <i>data monetisation on a data marketplace</i> ” <sup>113</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>114</sup> . E.g. data access may be provided through bilateral contracts <sup>115</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>116</sup> . The company wants to (b) <b>generate “revenue”</b> through data sharing <sup>117</sup> .	E.g. Dawex, Data Intelligence Hub.

<sup>108</sup> E. Scaria et al., p. 61

<sup>109</sup> E. Scaria et al., p. 61

<sup>110</sup> E. Scaria et al., p. 61

<sup>111</sup> E. Scaria et al., p. 61

<sup>112</sup> E. Scaria et al., p. 62. For further background information on data marketplaces see Florian Stahl et al.

<sup>113</sup> European Commission (2018): Towards a common European Data Space, p. 5

<sup>114</sup> European Commission (2018): Towards a common European Data Space, p. 5; E. Scaria et al., p. 62

<sup>115</sup> European Commission (2018): Towards a common European Data Space, p. 5

<sup>116</sup> European Commission (2018): Towards a common European Data Space, p. 5

<sup>117</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>118</sup>, “industrial data platforms”<sup>119</sup>, “direct data exchange between two companies”<sup>120</sup> and “data pooling”<sup>121</sup>.</p>	<p>An online platform shares data via a closed platform (e.g. independently, via a third party intermediary)<sup>122</sup>. For instance, a restricted group of users may voluntarily join an industrial data platform<sup>123</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>124</sup>.</p>	<p>E.g. Industrial data platforms: “Airbus and MAN created Skywise and RIO platform (respectively)”<sup>125</sup>.</p>

---

<sup>118</sup> European Commission (2018): Towards a common European Data Space, p. 5

<sup>119</sup> E. Scaria et al., p. 62

<sup>120</sup> H. Richter and P. R. Slowinski (2019)

<sup>121</sup> H. Richter and P. R. Slowinski (2019)

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

<sup>123</sup> E. Scaria et al., p. 62

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

<sup>125</sup> E. Scaria et al., p. 62



v.	Technical enablers <sup>126</sup> .	A technical enabler is a third party company that facilitates B2B data sharing between other companies through a “technical solution” <sup>127</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>128</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>129</sup> .	E.g. “API-AGRO, DKE-Data, Nallian and Sensative” <sup>130</sup> .
----	-------------------------------------	---	--	---

---

<sup>126</sup> E. Scaria et al., p. 63

<sup>127</sup> E. Scaria et al., p. 63

<sup>128</sup> E. Scaria et al., p. 63

<sup>129</sup> E. Scaria et al., p. 63

<sup>130</sup> E. Scaria et al., p. 63

## 4. Main policy issues

A number of policy questions have emerged without necessarily generating wide-ranging consensus on the answers. This is linked to the diversity of the role and nature of data in the online platform environment, in comparison with other sectors where data sharing has already been mandated or is being discussed (e.g. PSD2<sup>131</sup>, agriculture<sup>132</sup>, automotive, energy<sup>133</sup>). In addition, 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, also play a role. To exemplify this diversity, in what follows 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 necessarily on preliminary work<sup>134</sup>. 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<sup>135</sup>.

### Value of data: a premise

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 have become very skilled at designing strategies to collect data and 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

---

<sup>131</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>132</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>133</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>134</sup> Such as the report on “Competition policy for the digital era” and a literature review and formerly collected evidence sources.

<sup>135</sup> Other policy issues not addressed in this report relate to: i) Data governance for citizens; ii) Data sharing with competent public authorities or relevant third parties; and iii) Data for good initiatives.

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.<sup>136</sup>

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 as a source of market power*

The competitive relevance of data for all actors in the online platform economy, and hence the role of data as a foundation of market power, has been increasingly a topic of analysis for competition

---

<sup>136</sup> 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. Other studies do also estimate the consumer surplus value of these data-based services. Collis, Moehring and Sen (2020), 'Economic Value of Data: Quantification Using Online Experiments', NetInstitute Working Paper.

experts and not only. For instance, the expert group report on "Competition policy for the digital era"<sup>137</sup> recently pointed 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". Data reinforces each characteristic of platforms and thereby strengthens the tendency of market tipping and exposes the existence of economic power of platforms.

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>138</sup>.

Both, network effects and economies of scope in data generate a number of important challenges. First, 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. Second, 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>139</sup> Finally, 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>140</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

---

<sup>137</sup> Crémer, Montjoye and Schweitzer (2019), *Competition policy for the digital era*

<sup>138</sup> Zhu et al. (2017)

<sup>139</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 (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. 13, point I.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>140</sup> Stigler Center Report (2019)

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. Dominant platforms and powerful platform ecosystems that, thanks to their data related advantages, face no threat of entry from competitors, would have little incentive to innovate to increase consumer welfare. Hence, there is a risk that the potential benefit from data-driven innovations would remain under-exploited.

The economic power of large online platforms is therefore a key challenge for the economy as a whole. 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.

### Data sharing

Online platforms aggregate large amounts of data, which is at the very core of their business models<sup>141</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 these data collected by platforms

---

<sup>141</sup> See e.g. Shoshana Zuboff, *The age of surveillance capitalism*, Profile Books; Main edition (31 Jan. 2019).

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>142</sup>.

In the platform to business relationship, two main issues have emerged. First, although user-specific data are extensively shared by the platform with the business users, they lack access to more individualised data about their own operations or own customers/consumers or to a sufficient level of insight. Second, they require aggregated data on the entire market including also on competing business users' operations and their respective customers/consumers. For example, 33% of the 'heavy' users of online platforms responding to a survey<sup>143</sup> stated that they had experienced data access related issues in their trading practices with online platforms. A business user survey, carried out in 2019<sup>144</sup> showed that around 27% of surveyed enterprises disagreed, strongly disagreed or partly disagreed with the statement that the platforms share these platform-wide data with the business<sup>145</sup> users to a very limited extent.

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.

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, even though the trend is clearly for all companies to increasingly realize the potential of data. 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. 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

---

<sup>142</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).

<sup>143</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>144</sup> PPMI (2020), *Platform data access and secondary data sources*, Analytical Paper 1, Chapter 4.

<sup>145</sup> PPMI (2020), *Platform data access and secondary data sources*, Analytical Paper 1.



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

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. This also impulses the trend to “platformisation” in traditional industries like automotive or energy.

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>146</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. But, at the same time, more data sharing could also result in sharing commercially sensitive data with competitors, which platform users would be also reluctant to do.

### Data governance

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>147</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).

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 trade secrets considerations. Regulation (EU) 2019/1150 on promoting fairness and transparency for business users of online intermediation services (the so called “P2B Regulation”)<sup>148</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. The platform still maintains a privileged market overview because

---

<sup>146</sup> This could take the form of a ‘data cooperative’ or ‘data trust’; models are only emerging.

<sup>147</sup> Duch-Brown, Martens and Mueller-Langer (2017).

<sup>148</sup> OJ L186, p. 57, 11.7.2019.

it has access to data of all its business users resulting in information asymmetry to its own advantage (self-preferencing).<sup>149</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>150</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>151</sup>, research has evolved since, notably as regards IoT ecosystems<sup>152</sup>. Also, different EU and international reports have advocated for effective ex-ante regulatory measures<sup>153</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>154</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 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>155</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>156</sup> and data production often involves more than one entity. All these aspects make an allocation of rights to access, use and transfer between producers and users a very complex task.<sup>157</sup>

---

<sup>149</sup> Not using this valuable information could however result in welfare losses on the consumer side of the platform. Competition policy, and the use of data sharing as a means to overcome monopolistic behaviour, need to walk a fine line between avoiding the welfare losses from monopolistic behaviour and avoiding losses from not using the informational advantage. Whether data sharing among all parties is a solution is still very much debated. A recent economics paper discusses this dilemma - Kirpalani, Rishabh and Philippon, Thomas, Data Sharing and Market Power with Two-Sided Platforms (October 2020). NBER Working Paper No. w28023.

<sup>150</sup> Kerber (2017:16)

<sup>151</sup> "a comprehensive analysis of market failure problems is still missing", Kerber (2017:11)

<sup>152</sup> [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3445422](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3445422)

<sup>153</sup> Crémer, 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>154</sup> OECD (2015: 179-180)

<sup>155</sup> Duch-Brown et al (2017: 13)

<sup>156</sup> EPSC (2017: 5)

<sup>157</sup> HM Treasury (2018: 6); Datenethikkommission (2019: 104); Kerber (2017: 14)

Building upon the learnings of data sharing initiatives in different jurisdictions,<sup>158</sup> including initiatives funded by the European Commission such as the first EU-wide accelerator, Data Pitch<sup>159</sup>, an interdisciplinary debate around data governance is exploring various regulatory approaches for encouraging forms of sharing and granting access to data<sup>160</sup>. Zech<sup>161</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>162</sup>, he argues, would create incentives not only to produce but also to make available or reveal data<sup>163</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>164</sup>.

The discussion about other regulatory options is in full swing. In its data strategy of February 2020<sup>165</sup>, the European Commission emphasizes 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>166</sup>. Caution is also stressed against

---

<sup>158</sup> See e.g. GovLab. Data Collaboratives. Retrieved from: <http://datacollaboratives.org/>; Hub of All Things (HAT): <https://www.hubofallthings.com/>, Nesta

<sup>159</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>160</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, Alexsis 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>161</sup> Zech (2016)

<sup>162</sup> Zech (2016: 10)

<sup>163</sup> Zech (2016: 11)

<sup>164</sup> Duch-Brown et al (2017)

<sup>165</sup> [A European strategy for data](#), COM(2020) 66 final, 19.2.2020

<sup>166</sup> OECD (2019)

exclusively designing data sharing as a tool to hinder the consolidation of mature platforms and foster a rapid growth of new entrants.

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

## 5. Conclusions

Over the last two decades, 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>167</sup>.

While the European data economy is still evolving, 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 to help improve their services and their relationship with their customers, and to leverage data-driven network effects for generating additional revenues.

Research has shown that the proportion of data shared by companies usually depends on their business strategy. There are a number of different actors in the online ecosystem, which is cross-border and inter-sectorial by nature. These different actors have different interests in relation to data.

As a result, data may be shared against payment, or companies may preserve a certain discretionary autonomy with whom they want to share what data and under which conditions. Especially for smaller companies which do not naturally gain access to a critical mass of data in their activity, not having access to relevant data can have negative consequences such as missed business opportunities to improve relations with their customers, but also to innovate, scale up or streamline administrative processes.

---

<sup>167</sup> ACCC, 2019: 11.

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.

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 with particular focus on online platforms 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 the overall effect of potential data sharing obligations, 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. For instance, horizontal data sharing obligations could put at risk those new players’ existence.

The work for policy makers will therefore come down to answering specific questions, such as what kind of data should the regulator focus on when requiring platforms to share more data. Despite the incontestable appeal that simple rules across the board may have, when it comes to the data economy, such an approach will not work, and may risk doing more harm than good.

To achieve the goal of maximizing the benefits of data for all actors, and not only for a select few, regulators will have to delve into the details of what could appropriate thresholds be under which a platform is obliged to share data and how should that data be made available. To make things more complicated, the solution to such challenges might also differ depending on sectors or ecosystems where it would apply.

To support this challenging policy work ahead, we would recommend that the Observatory, in the next phase of its work on data (beyond the mandate of the expert group who has authored this report), focuses on some specific areas:

1. Continuous gathering of empirical evidence, to further deepen the understanding of data-driven ecosystems - 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. Further research on technical modalities for data collection and sharing (e.g. overview of the ecosystem of APIs and related emerging technology interfaces for data-sharing), so that

any future obligations or prohibitions on companies as regards data can take due account of technical possibilities and constraints.

3. Case studies to analyze the impact (in a sectoral approach if needed) of possible 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-Economics. 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, available at <https://cerre.eu/publications/big-data-and-competition-policy>

Richard Feasey, Alexandre de Streel, *Data sharing for digital markets contestability: towards a governance framework*, CERRE report September 2020, available at <https://cerre.eu/publications/data-sharing-digital-markets-competition-governance/>

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>

Drexl, 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-bd-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/ldecom/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=oe&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-beld-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.