Business user and third-party access to online platform data

Analytical paper 5
Authors: Vaida Gineikytė, Egidijus Barcevičius, Guoda Cibaitė
PPMI


The information and views set out in this report are those of the authors and do not necessarily reflect the official opinion of the Commission. The Commission does not guarantee the accuracy of the data included in this study. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use which may be made of the information contained therein.
# CONTENTS

Executive summary ........................................................................................................... 4

1. Introduction ...................................................................................................................... 8

2. Scope and methodology ................................................................................................. 10

3. Types of platform data used by other businesses .......................................................... 12
   3.1. Data for business decision-making and optimisation (downstream processes) .................. 13
   3.2. Data for the development of products and services (upstream processes) .................... 15
       3.2.1. Datasets for application development .................................................................. 15
       3.2.2. Data for developing “data-as-a-service” products for platform business users .......... 16

4. Data in e-commerce ........................................................................................................ 18
   4.1. Third-party sellers - the business users ..................................................................... 19
   4.2. Online marketplace optimisation tool providers ......................................................... 21
   4.3. Online advertising platforms – the gatekeepers ........................................................ 23

5. Data in the app store sector ............................................................................................. 24
   5.1. App developers ........................................................................................................ 25
   5.2. App store optimisation tool providers ..................................................................... 27
   5.3. Other platforms ........................................................................................................ 29

6. Data in the online travel services sector ........................................................................ 30
   6.1. Hoteliers and short-term accommodation providers .................................................... 32
   6.2. OTA optimisation tool and travel insight providers .................................................... 34
   6.3. Governments, cities, NGOs, researchers, journalists .................................................. 35
   6.4. Other platforms - the gatekeepers ............................................................................ 36

7. Analysis of incentives and constraints to data sharing ................................................... 39
   7.1. Economic barriers and incentives ............................................................................ 40
   7.2. Legal barriers .......................................................................................................... 42
   7.3. Technical barriers and incentives ............................................................................ 43

8. Possible solutions ............................................................................................................ 45
   8.1. Overcoming economic barriers ................................................................................ 45
       8.1.1. Binding regulation and mandated data sharing ..................................................... 47
       8.1.2. Self-regulation: access to data under FRAND .................................................... 49
   8.2. Limitations on data use imposed on vertically integrated platforms ......................... 53
   8.3. Overcoming technical barriers: data interoperability and portability ....................... 55
   8.4. Important considerations ......................................................................................... 56

9. Conclusions .................................................................................................................... 58

Annex 1. The case studies ................................................................................................... 63

Amazon ............................................................................................................................... 63
   Data provided to Amazon business users ....................................................................... 65
LIST OF FIGURES

Interview topics ......................................................... 133
List of interviewees ..................................................... 134

LIST OF FIGURES

Figure 1. Data value chain ................................................. 14
Figure 2. Ecosystem of data flows in e-commerce sector ............... 19
Figure 3. Data sharing ecosystem around the app stores ................. 25
Figure 4. Ecosystem of OTA data users ................................ 31
Figure 5. First page of Google travel search results ...................... 38
Figure 6. ARA Premium vs Brand Analytics ............................ 70
Figure 7. Website notice on Yasilv service discontinuation .............. 72
LIST OF TABLES

Table 1. Platforms selected for in-depth case studies ........................................................................ 10
Table 2. Prisoner’s dilemma ........................................................................................................ 46
Table 3. Types of data and analytics accessible by Amazon Marketplace sellers ......................... 69
Table 4. Types of data and analytics accessible to Allegro sellers .................................................. 75
Table 5. Types of data and analytics accessible by Aliexpress sellers .......................................... 79
Table 6. Types of data and analytics accessible by eBay sellers .................................................... 83
Table 7. Types of data and analytics accessible by Etsy sellers .................................................... 86
Table 8. Types of data and analytics accessible by Real.de sellers ............................................... 88
Table 9. Types of data and analytics accessible by Rakuten sellers .............................................. 91
Table 10. Types of data and analytics accessible by Zalando sellers ............................................ 94
Table 11. Types of data and analytics accessible by Google Play app developers on the platform .... 98
Table 12. Apple App Analytics metrics and dimensions ................................................................... 100
Table 13. Types of data and analytics accessible by Apple App Store developers on the platform ... 102
Table 14. Types of data and analytics accessible by properties listed on Booking.com ................... 107
Table 15. Types of data and analytics accessible by properties listed on Airbnb ......................... 112
Table 16. Types of data and analytics accessible by properties listed on Expedia ......................... 117
Table 17. Types of data and analytics accessible by properties listed on Tripadvisor ..................... 120
Table 18. Types of data and analytics accessible by eDreams business users ................................. 122

LIST OF BOXES

Box 1. OMOT role in supplying e-commerce business users with data ....................................... 21
Box 2. App Annie data collection .................................................................................................. 28
Box 3. Proposal for mandatory sharing of user information in data-driven markets by Prüfer (2020) ........................................................................................................ 48
Box 4. Data trusts ......................................................................................................................... 52
Box 5. Data collected by Amazon .................................................................................................. 63
Box 6. Amazon Brand Analytics, as accessed by a seller ............................................................... 66
Box 7. Data that Allegro collects .................................................................................................. 73
Box 8. Data provided by Allegro in Trade Analytics tool .............................................................. 74
Box 9. Data AliExpress collects ................................................................................................... 76
Box 10. Data provided to business users through the Data Latitude tool ...................................... 77
Box 11. Data eBay collects ........................................................................................................... 80
Box 12. Data analytics eBay provides ............................................................................................ 82
Box 13. Data Etsy collects or receives .......................................................................................... 84
Box 14. Data Real.de collects ....................................................................................................... 87
Box 15. Data Rakuten collects ....................................................................................................... 90
Box 16. Data Zalando collects ..................................................................................................... 92
Box 17. Data that Zalando provides to its sellers .......................................................................... 93
Box 18. Google Play Console: key metrics for app developers ..................................................... 96
Box 19. Apple App sales and trends reports ................................................................................... 101
Box 20. Analytics for Booking.com business users ....................................................................... 104
Box 21. Data and insights for Airbnb business users .................................................................... 109
Box 22. Analytics provided to Expedia’s business users ............................................................... 114
Executive summary

The two-sided markets operated by online platforms generate a lot of data about the players active on these markets. Such data is not only useful for the platforms, but also for other businesses to understand market dynamics, make better business decisions, or develop new products and innovate. The access to such data depends however on the terms and conditions set by online platforms. Decisions by online platforms on the scope and scale of data sharing have far-reaching effects on their business users and other players within the platform data ecosystem, and may help them succeed or put them at a disadvantage.

This analytical paper investigates the state of the art of data sharing by digital platforms with third parties. The analysis covers three sectors of the platform economy: e-commerce, online tourism services and app stores. It is based on a detailed research of secondary sources, 61 interview and 15 platform-specific case studies that include Amazon, AliExpress, eBay, Google Play, Apple App Store, Booking.com and others. Specifically, the paper was designed to answer the following questions:

— What data, collected and held by platforms, is important for their business users and other businesses active in their respective sectors?
— What kinds of data do platforms provide/open for each type of identified users? How and under what arrangements? What types of data, important to different users, do platforms refuse to share and on what grounds? How indispensable is this data?
— What are the incentives and constraints for platforms to share data?
— What are the possible solutions to address platform refusals to share data important to other users?

WHAT DATA IS IMPORTANT?

In this paper we differentiate between provided data that is inputted by the data subject; observed data, which is generated when data subjects use the platform, and inferred data, which platforms derive or calculate by examining various patterns of user behaviour. Further, data may come in different forms, that include raw data; data analytics (processed data, focusing on meaningful indicators) and insights (for example, suggestions on a specific pricing strategy, marketing approach). In all these categories, data can also be personal or non-personal.

Data is a key resource for businesses that use it for several purposes. In this paper we differentiate between (1) data used by businesses for optimisation of their internal processes, consumer relations and business decision making (downstream process), and (2) data used as an input to develop or improve data-based products or services (upstream process).

All data types collected by platforms are or could be important for business users for re-use. This includes data about transactions concerning own products and services, own clients/customers, and own business performance. Next, information concerning the broader market trends is also of key importance. It includes listings of other businesses, their customers, performance of different businesses in a specific market. Further, customer characteristics and customer profiles are of interest to all businesses, for example, behavioural data, such as browsing habits, search terms, purchasing decisions. The businesses using OTAs and e-commerce platforms underlined the importance of getting access to customer identification details e.g. for direct marketing. Finally, many companies, especially the smaller ones,
express their preference for data analytics and insights as they do not have sufficient infrastructure and skills to take advantage of raw data.

Some businesses also use platform data as an input to develop or improve data-based products or services (upstream process). In particular, the datasets of online platforms are of interest to two types of companies: app developers and data brokers or marketplace/app store optimisation companies. All types of data are pertinent to them, however they have a preference for granular and raw data that could be combined with other data sources and could be used to train algorithms, develop insights and provide value to their customers. More specifically, datasets and real-time data feeding into software and mobile applications can cover various areas and technologies, such as images for image recognition, audio files for speech recognition, weather or traffic data, health data, geolocation data and so on.

**WHAT KINDS OF DATA DO PLATFORMS PROVIDE AND WHAT DATA DO THEY REFUSE TO SHARE?**

Analysis carried out for this study shows that platforms provide data to their business users, which is sufficient to process transactions and manage their business. The businesses receive detailed 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. Overall, 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.

However, some data usually is not provided by the platforms, despite demand from their business users. Firstly, this concerns customer nominal data and contact details (especially pertinent in e-commerce and for OTAs). Secondly, the granularity of data concerning the customer profiles is also often considered insufficient by businesses. The key arguments for not providing customer data include personal data protection as well as risks to the platforms’ business model because businesses who can contact their customers directly may decide to bypass the platform. Further, some platforms argue that access to customer identification details may be used by businesses to send unwanted marketing messages. The majority of the interviewed businesses disagreed with these arguments. For example, they indicated that data protection may be managed at the level of individual businesses, provided customers are well informed and have an option to choose (opt-in) specific uses of their data.

Secondly, businesses demand more data about competing products and businesses on the platform. They also expressed a need for data about customer behaviour, such as search keywords, search volumes, buying patterns, responses to pricing signals. The platforms usually provide such data in a highly aggregated form and draw on it to develop analytics and insights that are offered or sold to business users. Nevertheless, many business users argue that such information is not sufficiently granular. Platforms argue that they cannot risk undermining the trust of their business users by sharing information that these users would not want to be made available to their competitors. Businesses that operate on the vertically integrated platforms (among online marketplaces, first and foremost, Amazon) also assume that the platform uses data from its marketplace to gain an unfair advantage over its own business users.

Our analysis also revealed power imbalances among platforms that are reflected in data sharing arrangements. Google and Facebook have the central position in online marketing and advertising, to the extent that they are unavoidable trading partners, including other platforms from the analysed sectors. This puts them in a position to determine the terms and conditions of data access and data reuse. Whereas Google and Facebook receive data from platforms concerning their listings, customers and business users, they do not share detailed data gained through the advertising activities. Further, some platforms also
signalled that data sharing arrangements put them at risk of being pushed out of the market by Google and Facebook that are developing their own business verticals in travel and e-commerce.

Finally, 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. The platforms that were analysed in this study argue 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. Platform-specific case studies also revealed several examples when decisions by online platforms (e.g. changing APIs, development of their own analytical services) undermined the business model of specific data brokers/ online optimisation tools providers.

Generally, all platforms claim that the only intended recipients for their data for re-use are their direct business users. Web-scraping is the main way to get access to platform data for all the other organisations interested in it. This is enabled by the fact that to generate transactions platforms must make a lot of information available for the customers on their websites.

**WHAT ARE THE INCENTIVES AND CONSTRAINTS?**

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. Generally, interviews with platforms revealed that they feel that they operate in an environment of legal uncertainty, which makes them reluctant to open more data. For example, they face different data protection regimes globally, as well as diverging interpretations of GDPR in EU member states. Further, whereas data sharing is usually considered as a measure to ameliorate power imbalances in the online platform economy, sharing seller-specific revenue information among sellers can be interpreted as providing a competitive advice under the national anti-trust law.

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 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. Next, when taking decisions on data sharing, platforms consider the global competition. For example, several platform interviewees pointed out that they detect abusive bots originating from China, crawling their pages or trying to use their APIs. Platforms see Chinese marketplaces as serious competitors that are not competing on a level playing field as they are in the position to disregard many regulations that European companies must comply with.

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.

**WHAT ARE THE POSSIBLE SOLUTIONS TO ENABLE MORE DATA SHARING?**

There is a clear public interest to encourage more data sharing, to the extent it could promote competition, offer more choices to businesses and their customers, foster innovation and help alleviate the market power of big online platforms. At the same time, the principles of personal data protection, business secrets’ and intellectual property protection should also be taken into consideration.

Various solutions have been put forward by various stakeholders that could potentially facilitate data sharing. They include both public-sector led initiatives, as well as market-based ones, focusing specifically on the incentives and constraints for data sharing stemming from the analysis. Public sector led solutions include mandated access; mandated interoperability and data portability; prohibition of certain business practices (for example, mandatory “walls” prohibiting vertically integrated platforms from sharing data between their marketplaces and product development / retail departments); and reversal of the burden of proof (i.e. platforms may be required to demonstrate that their data practices are beneficial for their users). Market-based or self-regulatory solutions considered include offering access to data based on FRAND (Fair, Reasonable, And Non-Discriminatory terms) principles; data pools or data trusts; as well as company-led incentives for interoperability and data portability.
1. Introduction

Data can serve various business purposes in the online platform economy and beyond, and be the basis for a wide variety of business models. At the very general level, distinction between two possible uses of data in business is important in this paper. First, companies can apply, process, and use data as an input to or specific part of their products or services they offer on the market (anything from software, data-as-service to management and consulting services). In the traditional production contexts, this could be seen as the upstream flow of inputs. Second, businesses can also analyse and use data to derive new knowledge used in business decision-making, business improvement and optimisation of internal processes. This, in turn, could be seen as involving data in a downstream process. In both cases, data is key to innovation and competition in contemporary dynamic online markets.

Especially within the past decade, higher data storage capacities, faster means of data transmission and more powerful computing performance have significantly advanced the possibilities of data-driven innovation. Big data is now used to innovate in almost all areas, from pop culture and entertainment to delivery of public services, not to mention the businesses active in the online economy.

Significant innovation relies on huge amounts of data and therefore on the availability of vast datasets. In this context, two-sided online platforms take up a very special position as compared to other economic actors. Through intermediating commercial transactions between businesses and consumers, platforms can oversee and record all the data on these transactions and the related actions of different parties. The ones generating most transactions, hold the most extensive datasets, not only enabling continuous innovation by the platform operator, but also relevant to other businesses in their ecosystems and beyond. For instance, Amazon is considered to hold the largest e-commerce datasets in the world, while Booking and Expedia are considered to be the largest depositories of tourism data in the world. Online platforms have largely built their success on continuously feeding insights from the newly collected data back into improvement of their intermediation and ancillary services.

Access to this data for the purposes of re-use for any third-party businesses is often very limited (as demonstrated by the Analytical paper on data and previous research). The business users of these platforms (who largely contribute to the creation of this data), as well as other companies operating outside the platforms, therefore, show increasing interest in accessing such data. The goals of innovation and improved competition are often cited as the key underlying reasons for more platform data openness and sharing.

Unsurprisingly, therefore, data regulation in general and the debate on access to data in particular have become the focus of discussions in both recent scholarship and current regulatory policies. As of today, there is still quite a lot of uncertainty regarding technical and legal access to data, as well as trust, applications and implications when it comes to the use of data by third parties.

---

1 For example, Netflix used Big Data to make strategic choice in developing one if its most popular TV series and instant success House of Cards.
2 For example, smart city mobility solutions.
This analytical paper investigates the state of the art of data sharing by digital platforms with third parties, including their current business users, as well as other businesses potentially interested in it. The analysis covers three sectors of the platform economy - e-commerce, online tourism services and app stores - as well as 15 specific platforms to provide insights into the following research questions:

— What data, collected and held by platforms, is important for their business users and other businesses active in their respective sectors?
— What kinds of data do platforms provide/open for each type of identified users? How and under what arrangements? What types of data, important to different users, do platforms refuse to share and on what grounds? And how indispensable is this data?
— What are the incentives and constraints for platforms to share data?
— What are the possible solutions to address platform refusals to share data important to other users?

To map the different actors using platform data in different sectors of platform economy, and the data flows between them, we apply the ecosystem view. The ecosystem metaphor for platform data users provides a way to understand the context of interdependencies and interrelationships, rather a sequence of one-on-one relationships between a platform and the various users of its data. Viewing platform data as part of ecosystems, allows to investigate the data sharing issues as those of networks of different claims to rights and or interests in the data. A variety of users may access - or wish to access - and use the data in different ways and by different means. At the same time, the data is generated and is about different actors - businesses, platforms and individual consumers - whose interests may be violated if the data on them ends up in the wrong hands.

In the following chapter we describe the methodology applied for the analysis, aimed at answering the research questions listed above. Chapter 3 then presents the typology of relevant data that we use further in sector- and platform-level case studies. Chapters 4, 5 and 6 cover the detailed synthetic sector level analyses on e-commerce, app stores and online travel services, respectively. In each of these chapters, we describe the main players and their interrelationships related to platform data sharing and exchanges (basically, data platform sharing ecosystems). The analysis largely builds on the 15 case studies on major two-sided online platforms, presented in Annex 1, as well as additional sources and insights. Chapter 7 then outlines the results of in-depth analysis, focused on the technical, economic and legal incentives and constraints that the overviewed platforms face when it comes to data sharing with other businesses. Chapter 8 presents the overview of the possible solutions to address the lack of platform data openness. Finally, Chapter 9 brings together and summarises the main findings.

5 This, however, should be distinguished from what is often referred to as “platform ecosystems”, presenting the view of different platform services, verticals and subsidiaries, and their interrelationships.
2. Scope and methodology

This paper investigates specific types of data that are generated on two-sided B2C online platforms in the process of facilitating transactions between their business users and consumers. The more specific data types covered in the paper are described in more detail in the following Chapter 3. Moreover, the paper looks only into the data shared or opened by platforms for the purposes of re-use by their business users and other third parties. This leaves the variety of data flows between platforms and their business partners or service providers (e.g., third-party technology solutions, payment processors, research, advertising or marketing service providers, etc.) outside the scope of this specific paper to the extent that these third parties are contractually not allowed to re-use platform data for their own business purposes.

Further, the analytical paper focuses on three broad sectors within the platform economy: e-commerce, dissemination of apps and software, and accommodation/hospitality. These were selected as the sectors and economic activities that are the most heavily and increasing intermediated by digital platforms. To illustrate, between 2014 and 2018 online hotel distribution in Europe grew by 46.7\(^7\). The growing e-commerce sector in Europe was estimated to reach sales of 621 billion euros in 2019\(^8\). Meanwhile, only a miniscule share of mobile application distribution takes place outside app stores.

For the in-depth analysis of the platform data collection and sharing practices, 15 platforms representing these three sectors were selected (see the table below).

### TABLE 1. PLATFORMS SELECTED FOR IN-DEPTH CASE STUDIES

<table>
<thead>
<tr>
<th>SECTOR</th>
<th>PLATFORM</th>
<th>VERTICAL INTEGRATION</th>
<th>GEOGRAPHICAL COVERAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-commerce</td>
<td>Amazon</td>
<td>Yes</td>
<td>Global</td>
</tr>
<tr>
<td></td>
<td>Allegro</td>
<td>No</td>
<td>Poland</td>
</tr>
<tr>
<td></td>
<td>Alibaba/Aliexpress</td>
<td>No</td>
<td>China and globally</td>
</tr>
<tr>
<td></td>
<td>eBay</td>
<td>No</td>
<td>Global</td>
</tr>
<tr>
<td></td>
<td>Etsy</td>
<td>No</td>
<td>Global</td>
</tr>
<tr>
<td></td>
<td>Real.de</td>
<td>Yes</td>
<td>Germany</td>
</tr>
<tr>
<td></td>
<td>Rakuten</td>
<td>No</td>
<td>Japan and globally</td>
</tr>
<tr>
<td></td>
<td>Zalando</td>
<td>Yes</td>
<td>Europe</td>
</tr>
<tr>
<td>App stores</td>
<td>Google Play</td>
<td>Yes</td>
<td>Global</td>
</tr>
<tr>
<td></td>
<td>Apple App Store</td>
<td>Yes</td>
<td>Global</td>
</tr>
<tr>
<td>Online travel agencies</td>
<td>Booking.com</td>
<td>No</td>
<td>Global</td>
</tr>
<tr>
<td></td>
<td>Airbnb</td>
<td>No</td>
<td>Global</td>
</tr>
<tr>
<td></td>
<td>Expedia</td>
<td>No</td>
<td>Global</td>
</tr>
<tr>
<td></td>
<td>TripAdvisor</td>
<td>No</td>
<td>Global</td>
</tr>
<tr>
<td></td>
<td>eDreams</td>
<td>No</td>
<td>Global</td>
</tr>
</tbody>
</table>


This selection allowed us to focus our analysis and get a comprehensive view of the broader data-sharing practices. First, it allows comparisons between vertically integrated and non-integrated platforms, which is an important distinction in research and policy discussions and may have very important effects on competition and innovation. The definition of vertical integration in this case is limited to the platforms engaging in activities on their marketplaces that are in direct competition with their business users (i.e., marketplaces that also have their retail activities, or app stores offering their first-party apps). Generally, most existing major app stores are vertically integrated (i.e. they offer first-party apps on their stores), while most OTAs are not (i.e. they do not run own hotels). Meanwhile, among major e-commerce marketplaces, we can find both vertically integrated and non-integrated. Our case selection therefore includes marketplaces to cover both models.

The selection covers the largest/ most successful platforms operating globally, as well as country-specific ones. This allows comparisons in several respects:

- Many businesses claim that they “must” be on the major global platforms to be successful in their respective businesses. Therefore, their case studies reveal the data sharing practices that many European platform business users are confronted with.
- Business user multi-homing is very common in the tourism/accommodation sector, so review of a variety of different OTAs allows to see the full picture of data access that they get from different intermediators.
- Data collection activities of smaller, local and non-integrated platforms are naturally of a smaller scope. Their comparison with the global marketplaces could provide important comparisons on the data access, uses, as well as underlying incentives and constraints to share data.

The evidence used in this analytical paper was collected through desk research and interviews. Given that most of the data on the research questions were collected from stakeholders involved directly in the platform economy and representing different interests, the research team triangulated and fact-checked the claims feeding into this report. Meanwhile, the positions of specific interest groups and parties are reported as such.

Desk research included the overview of the research literature (both grey and academic) on the issues of data in the platform economy. Given that the existing literature focusing on the specific questions of this analytical paper is rather scarce, the research team also reviewed various news articles, blogs, and internet forums, which also served as a primary data source. The researchers used this information for two purposes. First, as a source of factual information, after fact-checking and triangulating with other sources (e.g., we investigated whether some types of data are provided by platform or not, by triangulating business user impressions, platforms’ own presentation of their policies and interview data). Second, as a primary source of data representing positions and perceptions of different actors in the platform economy. The full list of reviewed sources is provided at the end of this report.

Furthermore, the analysis presented in this paper is based on 24 new interviews conducted for this analytical paper and **61 interviews in total** (including those conducted for other analytical papers, which presented relevant insights for this paper as well) with online platforms, their business users, associations representing each of these groups, and experts working on issues of data access and regulation. The lists of main interview topics and interviewees is provided in at the end of this document. The interview questionnaires focused on the research questions, and were customised on a case-by-case basis, taking into account the interviewee affiliations and knowledge or experience on the topics.

The data collected first fed into the detailed platform case studies, presented in Annex 1. The case studies were developed following a unified template, building on the conceptualisation of platform data and its users presented in Chapter 3. Then, the synthetic analysis of case study data, as well as the investigation of additional sources, resulted into sector level and synthetic analyses, presented in this paper.
3. Types of platform data used by other businesses

Broadly speaking, multisided online platforms collect, control and use three broad types of data, classified by its origin:

- **Data actively and knowingly provided** by the data subject (for example, mailing address, user name, age, company information, product or service reviews, etc.) at the time of registration, transaction, communication and other.

- **Observed data** “provided” by the data subject by virtue of the use of the service or the device. They may for example include a person’s device information, search history, traffic data, pages or links clicked on, and location data. This kind of data is sometimes referred to as “digital exhaust” or “data exhaust”. It may also include other raw data such as the heartbeat tracked by fitness or health trackers, biometrics from smart phones, voice recordings from smart home assistants or data from other IoT devices.

- **Inferred and derived data** created by the data controller on the basis of the data provided by and gathered from the data subject. In this case data subjects can be both, business users and consumers using the platform. Inferred (also called “implicit”) data includes the characteristics assigned to people based on their activities and behaviours, often around content consumption such as online searches, artists or playlist followings, public information shared on social media, location history or recent purchases, subscriptions and so on. Inferred data is derived from and can be further matched with provided and observed data to enrich a customer (or business) profile already created based on other users’ information. Such behavioural data enables personalised insight to customer lifestyle interests, political and religious inclination or buying habits, which can in turn be of a very high value for marketers.

Platforms have largely built their services and **business models** around the continuous use of this data, for example, by feeding the insights based on it back to the intermediation service improvement, developing new products of services, expansion into new markets by developing new business verticals, and monetisation of behavioural data though personalised advertising. As the analysis will show further, this is closely related to the incentives and constraints of platform data sharing with third parties.

Meanwhile, such data controlled by platforms is as important to other businesses (some of which largely contribute to its creation) as they can use it:

1) to make better business decisions, identify market trends and niches, or optimise business processes (downstream process), or

2) to develop (with the goal to offer them on the market) products and services, which require data as an integral part (upstream process).

---

9 [https://www.techopedia.com/definition/30319/data-exhaust](https://www.techopedia.com/definition/30319/data-exhaust)

We further examine the types and classifications of data demanded for these two distinct uses by the current platform business users and other actors.

### 3.1. Data for business decision-making and optimisation (downstream processes)

Some earlier studies, as well as insights collected in preparation of this analytical paper allow to crystallize several more relevant dimensions of what is broadly referred to as **data relevant for business decision-making** (speaking specifically about platform business users). To begin with, a study by VVA distinguishes between six categories of data generated by online platform users, to facilitate an operational analysis of data sharing practices in the P2B relationships:\[11\]

- **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). This includes data which allows to identify the customer across different platforms, e.g., marketplaces and social media.
- **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 products/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 (clicks, browsing history, other products or services purchased on the platform, provenance on the Internet, 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”).

While identification details are valuable for directly targeting or reaching out to specific individuals or other businesses, other data types feed into designing the appropriate ways to implement this targeting, and/or improve the business outcomes more generally.

Then, especially when talking about data relevant for the current platform business users, the following distinction is important:

- **Data on the business’s own** listings, 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.
- **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 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\textsuperscript{12}.

As the earlier analytical paper\textsuperscript{13} on platform data access showed, while access to the first category of data is largely not problematic, lack of access to the second category often becomes an issue for business users - especially those operating on vertically integrated platforms. Most discussions about the lack of platform data access for downstream processes revolve mostly around this information about the market trends, competition and comparative business or product performance, market niche identification and so on - accessible by platforms (and in some cases by their verticals competing in the same markets), but not by other market players.

Furthermore, several levels of data and its use and applications exist, related to the data value chain\textsuperscript{14} (see the figure below). Based on the data value chain framework, data follows a consequential approach, and at each step its different features are important. Looking into value as the key feature, most of it is derived from data once it is processed and used for business decisions (in the figure below, the “++” signs graduate the importance of data characteristics at each step). The typology of six categories of data presented above already integrates this view, introducing a hierarchy of data to some extent (from identification details being at the bottom, and market analyses at the top). The platforms create this value by collecting data on various actors in a centralised way, as well as aggregating data and drawing insights. Meanwhile, the business users have a restricted access to it, often limited to data on their own business, rather than broader overview of the marketplace.

\textbf{FIGURE 1. DATA VALUE CHAIN}

\includegraphics[width=\textwidth]{data_value_chain.png}


As discussed further, for upstream processes - in which data is seen as an input (comparable to a raw material in traditional production processes) - the volume, velocity and variety of data are the most important. However, for the downstream processes, data is as useful, monetizable or needed, as it can be effectively used in making business decisions. As the figure above illustrates, therefore, its value increases with the level of processing. Analytically, therefore, data can be divided into types related to its aggregation:

- First, it is \textit{raw/ unaggregated data} collected by the platforms, business users or third parties. It can be provided to platforms by consumers or business users, or observed.
- The second type is often referred to as \textit{analytics} - processed data, focusing on meaningful indicators. This type encompasses various levels of data aggregation, granularity and processing. Inferred data also falls under this category, as it requires some automated data processing combining of data sources and data points.


\textsuperscript{13} Forthcoming.

The third level is insights, directly applicable in business processes. They build on data analytics to provide a normative dimension: what a business should do based on the evidence available.

In addition to these classifications, various sources tend to often distinguish between personal and non-personal data held by platforms. This is important from a legal perspective, as this distinction will have repercussions on the legal regime applicable for data access and data sharing. However, from the business perspective, this distinction is difficult to make (see more details in Section 7.2). For example, most of data types data can be depersonalised, and much of it can be re-personalised. We therefore further use the distinction between personal and non-personal data in cases when this is relevant for platforms and other companies in data exchanges.

### 3.2. Data for the development of products and services (upstream processes)

High quality data is the main input for developing applications of Artificial Intelligence (AI), software, mobile applications and data products (data-as-a-service). As it is shown further in the analysis, two types of companies mostly benefit from datasets acquired from digital platforms: app developers, which are also a group of platform (i.e. app store) business users; and marketplace or app store optimisation companies, as data on the activities taking place on the platforms is their main offering on the market. They are presented in more detail in the following sections.

#### 3.2.1. Datasets for application development

Datasets and real-time data feeding into software and mobile applications can cover various areas and technologies, such as images for image recognition, audio files for speech recognition, weather or traffic data, health data, geolocation data and so on. It is usually generated and collected not through platform intermediation activities. However, the platform operators are usually the entities who hold the most extensive data in these markets (for example, in case of Facebook - pictures to train facial recognition software; in case of Google - maps and traffic data, Amazon - voice recognition, and so on).

The specific datasets that are needed depend a lot on the type and specific characteristics of an app or its specific functionalities. According to an interviewed developer’s association, whenever there is a demand for specific data, an ecosystem develops around it. Data companies active in these markets build sophisticated tools that allow their customers (i.e. app developers) to access the data in a controlled way, providing the specific data that developers need and pay for. For example, numerous companies exist in the geolocation data segment, engaged in collecting and selling this data. Many app developers subscribe to their geotechnical datasets. Various API marketplaces emerge to help developers to easily reach the data.

Players like Amazon and Google also act as data providers and are, for example, launching speech recognition or translation engines as a service for app developers. Generally, this data industry is maturing, providing app developers with convenient access to multiple sources of data and allowing them to avoid digital hoarding, and get a convenient access to the specific data they need, when they need it. While proprietary real-time data services provided by companies such as Google are quite expensive

---


16 For example, https://rapidapi.com/
for the smaller developers, a lot of open source substitutes also exist\(^\text{17}\), which are widely used by app developers.

### 3.2.2. Data for developing “data-as-a-service” products for platform business users

Companies providing marketplace optimisation services for platform business users are present in all the sectors of the platform economy overviewed in the paper. They are generally interested in the types of platform data necessary for the decision-making of businesses active on the platforms (i.e. for downstream business user processes, as described above). The difference is in how they use it, that is as an input in the data products that they sell to platform business users.

Especially interesting are the ways and sources that they collect, infer and aggregate the online platform data, which are out of reach for individual platform business users. Research implemented for the previous Analytical paper\(^\text{19}\) allowed to distinguish several key sources of platform data, that these entities build their services and products on.

**Web scraping.** Online platforms host a significant amount of publicly available data on their websites, which is important resource for optimisation tool providers serving the platform business users. Although with proper tools it is available for anyone to collect, doing this efficiently requires sophisticated instruments and know-how, which most business users do not invest into. Nonetheless, a large number of third-party data companies engage in scraping of platform websites to provide these data and insights based on it. For example, Amazon representatives revealed in an interview that around 35% of the traffic on Amazon’s stores comes from web scraping tools. The online retailers and marketplace optimisation companies contribute significantly to this. As a blogger and speaker on Amazon selling explained, the platform’s approach to the data companies is “to let them ride” by leaving “the backdoor open”, although this practice is not in line with the platform’s formal terms and conditions.

There is also whole data scraping industry connected to the two-sided platform markets, not limited to the optimisation tool providers. Many companies, which specialise in technological solutions for data scraping exist, and there is a whole ecosystem of other companies in the supply chain. To give a couple of examples, a company named ProxyCrawl\(^\text{20}\) offers anonymous crawling and scraping tools and services; MyDataProvider\(^\text{21}\), among other things, specialises in scraping e-commerce marketplaces, such as Amazon, AliExpress, Airbnb and eBay. Also, a number of companies offer multiple Internet Protocol addresses (IPs) for cases when the bots scraping platforms are identified and their IPs are banned. Some of these companies reportedly buy thousands of servers and IPs in the US upon agreements with telecom providers and use these IPs to scrape online marketplace data in large volumes and overcome the platform measures aimed at limiting these activities.

**Crowdsourcing business user account data.** As described in Chapter 4, each seller on e-commerce marketplaces - as well as business users in other sectors - get access to their own business insights, as well as some insights concerning their competitors and the broader market overviews. Some third-party data providers ask online sellers to share such marketplace information, and then link this data from thousands of business users and scraped data to draw market insights. For example, Jungle Scout collects data from

---


\(^{18}\) For example, http://opentrack.io/

\(^{19}\) Analytical paper #1: Platform data access and secondary data sources.

\(^{20}\) https://proxycrawl.com/about-us

\(^{21}\) https://mydataprovider.com/solutions/web-scraping/ecommerce/
a large number of sellers (over 225,000) who have opted in to share their sales information with it\textsuperscript{22}. When merged with the data gathered by scraping the platform’s front-end (e.g., Best Seller rank on Amazon), this can be used for quite precise estimations and extrapolations. Similarly, if AirDNA users wish to receive performance analytics, they are asked to upload their Airbnb host IDs. After doing this, AirDNA enables the hosts to see their performance trends, comparative and financial analysis in comparison to the overall marketplace listings. Other data brokers and analytics companies in the app sector, such as Beyond Pricing, Apptopia (with 300,000 developers opted-in), Wheelhouse, AppAnnie (with a million developers opted-in) and others, also use such technique as an important data source.

**Processing and analysing data provided by platforms.** Some platforms tend to provide some raw data themselves, e.g. through public APIs. As many business users do not have know-how and infrastructure to work with such data, third-party data providers step in to fill this gap. For example, Terapeak provided such service for eBay sellers and continues to do so even after being bought by eBay.

Overall, the commercial relationship between platforms and optimisation tool providers vary from case to case. Most platforms claim that they are not in data sharing relationship with optimisation tool providers, so they could not comment on the validity of data that the third-party data companies provide. Several interviewed data companies, in turn, mentioned that they have good relationships with the platforms they focus on. Most platforms seem to have a certain degree of tolerance for web scraping activities, although they put technical measures in place to limit such data collection. The reviewed online sources also show that some platforms invest in preventing such data collection for commercial purposes. However, that does not prevent data companies from collecting data this way. As the HiQ vs LinkedIn court case on web scraping in the US illustrates, it is difficult for platforms to enforce the prohibition of automated collection of their publicly accessible data\textsuperscript{23}.

Nonetheless, the business model which relies on the data displayed publicly on platform websites or public APIs is quite risky. First, the business model fails if the platform starts providing the data or analytics offered by the data brokers. For instance, Allegro, a Polish e-commerce platform shares most of the data on the activities on their platform via an API. According to an interviewed platform representative, it has been long used by an insight provider, but recently the platform itself made significant investments into analytics, to develop new data products (Allegro Statistics) that are now provided to its sellers. A different example is Amazon, which 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. Some of the optimisation providers ultimately adapt to such changes by leveraging precise historical data from the API together with sophisticated means of forecasting and estimating\textsuperscript{24}. However, their insights product does not have the same 100% level of accuracy anymore, as it now relies on inferred and not the action data.


4. Data in e-commerce

The European e-commerce sector shows continuous growth. Its turnover was forecasted to hit 621 billion euros in 2019\(^{25}\). A Mastercard survey from 2017 shows that one in four Europeans who have internet access have shopped online at least once a week, while over 60% shopped online once a month and 6% even bought products or services via the internet every day\(^{26}\). Most of the e-commerce revenues, however, is generated in Western Europe, which accounts for approximately 66% of total European online retail turnover. Southern Europe, Northern Europe, Central Europe and Eastern Europe show a much lower share of European e-commerce with 14%, 9%, 6% and 4%, respectively.

Online platforms enabling B2C e-commerce transactions have significant shares in this market. While many European e-commerce marketplaces are present in national e-commerce markets of the Member States, American platforms have significant influence on local e-commerce industries in Europe as well. For example, Amazon, was the most-visited online marketplace in Europe in 2018. Such market status allows the e-commerce giants to compile the largest and most comprehensive datasets on European consumers as businesses.

What is more, some key players in the market, including Amazon, have a vertically integrated business model and act both as retailers and marketplaces enabling third-party sellers to offer their products to international markets. This raises additional issues related to data collection, use, as well as data sharing with the third-party merchants.

Platforms, compared to other actors in their data ecosystems, have the greatest possibilities to collect data on individual consumers. As the case studies on Amazon, eBay, Allegro, AliExpress, Etsy, Real.de, Rakuten and Zalando show, several distinct types of actors are interested in e-commerce platform data, although they have very different access to it (see the figure below). E-commerce marketplaces most actively share their data with business users, which are the primary intended recipients of platform data. Other entities, such as online marketplace optimisation tools providers, usually pool data from multiple sources, including publicly available data, crowdsourced business user account data, data provided by platforms through APIs and web scraping platform websites, as described above. Finally, online advertising platforms (i.e. Google and Facebook), which are regarded as potential or de facto competitors by the interviewed marketplaces, gain marketplace data as part of marketing and advertising service provision. In the following sections, we review the main types of platform data users in more detail.

---


\(^{26}\) [https://ecommercenews.eu/ecommerce-in-europe/](https://ecommercenews.eu/ecommerce-in-europe/)
FIGURE 2. ECOSYSTEM OF DATA FLOWS IN E-COMMERCE SECTOR

It is important to note that e-commerce platforms also hand over data to companies acting as service providers to them, such as marketing and analytics companies, shipping, carrier and delivery providers, payment providers, personal email communication providers, cloud infrastructure and data storage providers etc. This is especially prevalent among e-commerce platforms which have a ‘pure marketplace’ model (i.e. are not vertically integrated and do not offer such services as warehousing or logistics), such as eBay or Allegro. However, in this case, data is provided for the purposes of delivering services to the marketplace itself and not for re-use.

4.1. Third-party sellers - the business users

The third-party sellers or merchants active on the e-commerce marketplaces are the main intended target group for which the platforms provide the data for the purposes of re-use. Platforms display the data, analytics and insights to their sellers though dashboard accessible on seller accounts, in reports, or through specialised APIs.
Interviews with e-commerce sellers and desk research show that the distinction between the data on the seller’s own activities and overall activities on the platform is very relevant when talking about the data provided and unmet data demand. Regarding the first category, the sellers get most of the following data from the e-commerce platforms:

- Customer data, sometimes including phone number and address that are crucial to safely deliver the purchased goods. However, email addresses are provided very rarely, and phone numbers are also excluded by some platforms (e.g., Etsy).
- Financial data to understand business performance and margins.
- Listings, sales and product data to manage stocks.
- Advertising data, attribution rates, return on investment, to make better informed and targeted advertising decisions - for business users that subscribe to platform advertising products.
- Customer interaction with their listings, including how do customers reach their listing pages, click-through rates, bounce rates, conversion, reviewer characteristics. This is provided at various but generally low level of granularity and comprehensiveness.

What sellers commonly identified as an unmet data need was the market analytics allowing them to better understand customer behaviour, market trends and competition (i.e., the data on overall platform activities). More specifically, to be successful in online marketplaces, sellers need several types of data - all concerning a broader perspective than their own business operations and transactions:

- Data to understand the success of other players on the marketplace so that they could understand better the competition. For example, how many competing sellers there are for the specific range of product, what products they sell, how many and what reviews they get, what are their rankings, and so on. This includes competitor’s sales data. It helps sellers better understand the demand and its seasonality, make more precise marketing decisions, ensure that they are using the right keywords, choose the right pricing, optimize their advertising campaigns and track own performance and profits in a more informed way. Such data could also help to uncover products with large sales potential. For large companies, and especially luxury brands, sales data and details of counterfeit products is also important.
- Price data of products similar to theirs, to develop (dynamic) pricing strategies. For example, Amazon’s algorithm chooses which seller wins that default position — the Buy Box. While the exact formulas used to pick the winner are not known, Amazon’s website advises sellers that they can increase their chances by having low prices, having items in stock, offering free shipping and getting excellent customer service ratings. To optimize their chances, many sellers start using algorithmic software to constantly change prices to adapt to competitors’ moves.27
- Customer behaviour, especially search keywords and search volume data, to optimize their stores for search, as well as identify new products to improve, produce and sell28. Understanding their reviews and shopping habits, responses to texts, images and videos is also crucial.

The seller interest in this data is supported by previous research, also showing that companies using online marketplaces are often unsatisfied with the levels of access to such data they receive from platforms. For example, around two thirds of sellers who responded to the VVA survey disagreed with the statement that the information they receive through the online marketplace about the behaviour and preferences of customers is useful for the development or improvement of their products or

---

services\(^{29}\). Our business user survey showed that 52% of enterprises whose key activity on platforms is selling goods\(^{30}\), cannot access some types of data held by the platform, which are essential to their business. The issue of access to the platform-generated data is especially relevant in vertically integrated e-commerce platforms. Without data access, third-party sellers cannot viably compete with the marketplace controller retail verticals, as they get unique insight into the market. Nonetheless, as it is explained below in Box 1, optimisation companies - another major user of platform data - fill this gap to some extent.

Generally, as illustrated by the case studies (see Annex 1), all the overviewed e-commerce platforms provide their sellers with the necessary data to conduct the transactions, as well as some business analytics. The volume and granularity of this data, however, varies. Although significant differences were not identified based on whether the platforms are vertically integrated or not, generally the larger international platforms tend to provide more business analytics than platforms operating within national markets.

4.2. Online marketplace optimisation tool providers

There is an entire industry of services clustered around e-commerce marketplaces – companies offering anything from accounting, copywriting, advertising, seller overdraft (credit providers) to the development of sophisticated data-powered tools for sellers. In this paper, we focused specifically on third-party data companies, which are also known as “online marketplace optimisation tools” (OMOTs) in the context of e-commerce sector, because marketplace data is at the core of their business model. By developing products based on it, these companies address the unmet data demands by the marketplace sellers (see the box below). While some OMOTs are very focused on specific marketplaces, others collect data and develop insight products on a number of big players in the e-commerce sector or general sector overviews.

BOX 1. OMOT Role in supplying e-commerce business users with data

There is a widespread and growing demand for OMOT services by e-commerce sellers. Some reports forecast that the global online marketplace optimisation tools market will grow at a compound annual growth rate of 12% during the period 2019-2026\(^{31}\). Based on desk research and interview material, their use is widespread among marketplace business users and, as some of these companies claim, competing marketplaces themselves. Among the key players in this industry are JungleScout, Sellics, Feedvisor, AMZ.One, A2X, Helium 10, repriceexpress, SellerApp, SellerExpress, Viral Launch, ByteStand, eComSpy, Shopkeeper, Teikametrics, Marmalead.com, Algopix and Ahrefs.

Marketplace optimisation companies specialising in e-commerce, mostly provide insights on business demographics, market, consumer, product trends, and similar data. Most of their services are delivered as standalone software products, or add-ons for Amazon and other sites\(^{32}\). Some companies, such as for instance, Teikametrics for Amazon business users, provide sellers with data-powered tools to build successful brands by discovering new keyword


\(^{30}\) That is, 86 out of 165 businesses that use platforms to sell goods only but do not engage in selling services, listing accommodation or distributing apps.


\(^{32}\) Ibid.
opportunities, automating keyword bids or eliminating wasteful advertising which does not generate the satisfactory levels of conversion\(^{33}\). The analytics and insights that they offer mostly focus on individual seller’s business and advertising performance, based on the data the seller provides (which, in turn, is often generated and provided by platforms). Fewer marketplace optimisation tools provide transaction information and customer demographics, and basically none offer detailed customer data.

A significant number of data providers offer these services for the businesses active on each major platform. For instance, Jungle Scout and similar providers, such as Helium10, AMZScout and Unicorn Smasher, as several interviewed Amazon sellers explained, supply them with comprehensive market insights and competitor overviews that the platform has itself, but does not share. They offer such analytics products as data on best-selling listings, monthly revenues of other sellers, search volumes and search trends. Although these marketplace optimisation tools provide estimations and extrapolations based on what data they could gather rather than exact information, the sellers claim this data is very valuable given the Amazon’s reluctance to provide such insights for the sellers. In fact, some of the seller-oriented materials online warn them that if they do not want their competitors to see their revenue, they should stay away from Amazon - given that OMOTs focusing on this platform can collect a lot of detailed data on how the individual business users are doing. OMOT targeting this platform allow sellers to acquire substantial data on the competitors, and the estimations are said to be “spookily accurate” \(^{34}\). As an interviewed seller explained, if your competitors use these insights, you must use them too to be successful; data disadvantage can be very costly.

Also, Amazon, eBay and Etsy sellers note that compared to the analytics and insights provided by platforms themselves, third-party providers are faster and more comprehensive, especially useful if the sellers need to quickly understand how their business are doing. However, their data is not always precise, and in some cases can give only a very rough idea on how things actually are. Moreover, they may not always know how marketplace algorithms change to be able to swiftly adjust.

Marketplaces generally claim not to have a direct relationship with OMOT providers when it comes to data sharing since they are seen only as secondary (non-intended) users of marketplace data. These leaves OMOTs with several ways to access platform data from other sources:

- Publicly available reporting data, such as platform operator annual reports and other materials prepared and shares by platforms (e.g., Amazon’s Most Wished For\(^ {35}\) or Movers & Shakers reports\(^ {36}\), showing the products with the highest increase in rankings in the past 24 hours, aimed at introducing customers with trendy items).
- Web scraping (platform websites). Marketplaces make a lot of information on individual listings visible for the consumers to help them make the purchase choices, including sellers, prices, item descriptions, ratings, reviews, relative popularity and so on. Once scraped, merged and systemised, this data can allow a lot of further inferences into the market.
- Open data provided by platforms through APIs. While APIs are intended for business users and developers, they are accessible to other entities (only intrusive uses of APIs and web scraping are forbidden by Terms and Conditions of platforms).
- Crowdsourcing business user account data, as described in section 3.2.2.

The interviewed platforms reported certain degree of tolerance towards third party actors, as long as they do not interfere with their business. Firstly, because scraping bots are difficult to prevent in practice.


\(^{35}\) Available at https://www.amazon.com/gp/most-wished-for

\(^{36}\) Available at https://www.amazon.com/gp/movers-and-shakers
Secondly, because such tools can be beneficial for sellers and consumers. While sellers use OMOT’s services to make better informed business decisions (and ultimately generate more revenues for the platforms on which they operate), consumers benefit from the innovative products, price adjustments and better shopping experiences. Nonetheless reliance on such data collection leave the OMOTs highly dependent on data display and sharing policies of platforms, as described in Section 3.2.2.

4.3. Online advertising platforms – the gatekeepers

Some of the interviewed marketplaces raised concerns over data sharing with what they saw as dominant players in the advertising/ marketing sector: Google and Facebook. The interviewed marketplaces use Google and Facebook for marketing and advertising purposes. According to some of them, they see these players as hard to avoid since they are the main channels allowing to target and reach customers. Several interviewed marketplaces expressed concerns over this dependency on these platforms, which they usually call the “advertising partners”. They are wary of being “disintermediated” as there have been instances when data from marketplaces is used by these players to enter or expand into related supply or customer markets.

The big advertising platforms seem to exploit this dependency of other players by setting up terms and conditions unilaterally in contracts of adhesion with standard conditions. It has been described by several marketplaces as a “take it or leave it” situation. “Leave it” does not seem to be an option if a company wants to stay in e-commerce business - big advertising platforms play a gatekeeper role to the market due to their outreach and user/viewer base. “Take it”, however, means that marketplaces share with these advertising partners a lot of data on their listings, customers and business users. While marketplaces certainly do not intend this data to be shared for re-use by Google or Facebook, their contracts - prepared unilaterally by these advertising platforms - do not preclude this as a possibility. Moreover, according to several interviewed marketplaces, both Google and Facebook do not, in return, share detailed data gained through advertising activities.

This demonstrates a potential conflict of interest, if a dominant platform appropriates data for developing products that are competing with the marketplace products, based on the data that the marketplace provided. This also prevents optimisation of resources invested by the marketplaces in advertising or marketing, as it does not offer insights into the services received (access to such data could lead to better decision-making). Particularly smaller, not vertically integrated marketplaces seem to be mostly affected by these practices and have expressed a desire for more fairness and transparency in this relationship.

---

37 Please note that this information is not presented under platform case studies as the platforms that pointed out these issues did not agree to be identified.

38 For example, when platforms sign contracts for services requiring data transfer with other third-party service providers (e.g., payment providers, consultancies, etc.), they include a clause that this data may not be used for any other purposes than the provision of that service. This is not the case with Google and Facebook, making it legal for these giants to use the data of other platforms however they wish to.
5. Data in the app store sector

Among the sectors of platform economy overviewed in this paper, app stores market can be characterized by the largest concentration, with two key players listing the absolute majority of world’s mobile applications. Apple App Store is the only app store for iOS devices, and Google Play Store is the absolute leader among the Android OS app stores, while over 99% of all smartphones in the world run on either of these two operating systems (Android with over 85% share, and iOS with over 13%)39.

Given the popularity of both iOS and Android devices in the EU, for developers aiming to reach every European consumer with a smartphone, it is important to have an app present in both app stores. These, in turn, have the power to determine what functionalities are available to app providers to utilize, what type of content or services they can offer in their apps, and what data app developers can collect on their users.

It is important to emphasize that each mobile app itself operates as a tool for data collection as consumers interact with it. As Apple App Store and Google Play Store enable third-party software developers to develop and offer apps for Android and iOS, they also act as enablers and gatekeepers of their direct data collection from the consumers. Other online platforms - including those overviewed in this paper, as well as Apple’s and Google’s first-party software - also run through apps listed on the app stores. In fact, the launch of smartphones since mid-2000s has led, among other things, to the rise and massive growth of online platforms that facilitate and organise online interactions between users and suppliers.

Nonetheless, Google and Apple, as well as their respective app stores, have very different business models and thus different incentives related to data collection and sharing within their ecosystems and beyond. Generally, Apple’s mode of operation is quite commonly known as a “walled garden”40, as Apple is primarily a producer of electronic devices running on its own operating systems, sales of which generate most of the company’s income. Meanwhile, Google monetises on data products and applications, primarily in behavioural advertising, while Android OS runs on devices manufactured by thousands of companies. Subsequently, Apple highlights the importance of privacy and security as part of the company’s DNA, while Google promotes their more open ecosystem with lower priced and even free services41, but focuses on data extraction form its users. Nevertheless, both Apple and Google collect a lot of consumer data and business data, and not only through the app stores, but also through the ecosystems of their different products and services more broadly (e.g., first-party apps, devices, etc.).

The primary third-party users of data generated through app store multi-sided activities are app developers. The infrastructures collecting and processing these data on both app stores were developed with the business users in mind. In contrast to business users in other sectors, app developers themselves also have the capacity to collect a lot of data (including behaviour data) on the consumers, as they interact with the apps. Both app stores claimed they do not have any arrangements of data access for other commercial entities. However, our research showed that the ecosystems of app store data access and

39 https://www.idc.com/promo/smartphone-market-share/os; these number vary in other sources (e.g., over 74% for Android and almost 25% for iOS, according to Statista: https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/). However, the total share of the two stores remains over 99%.
exchanges involve more types of companies (see the figure below). In the following sections we investigate the main actors in the app store data ecosystems in more detail.

5.1. **App developers**

App developers are a specific group of platform business users, as data is not only needed to understand their business performance - it is the **key component of the products they offer** (as described in Section 3.2.1.) and feeds in all of their upstream processes. Moreover, app developers also have broad possibilities to collect data themselves using their own software, both for upstream and downstream purposes. Apps can interact with the user’s device to retrieve and share such information as usage data, photos, biometric data, financial information, contact details, login credentials, correspondence, location, IP number, etc. Some apps can also activate the device’s functions (e.g. camera and microphone) to collect additional personal data. All this data collection depends on end-user consent that the developers can acquire directly from the consumers (which is usually the precondition for consumers to be able to use the app). In fact, interviewed developer representatives confirm that data collected directly from end-user’s devices is of key importance for their activities. On the negative side, however, the extent of their data collection activities sometimes is not entirely compliant with the data protection rules and privacy (notably as
regards data minimisation). Very often, end-users do not fully understand the permissions that they grant to apps that they use. Moreover, certain app developers may use the data they collect through their applications in intrusive, unethical or illegal ways. In this context, the public discourse often presents app stores as the gatekeepers protecting end-user’s privacy from app developers, but sometimes failing at this job.

Before the implementation of recent data protection regulations and most notably the GDPR, app developers had much broader possibilities to collect data allowing to personally identify the end-users, their devices and other personal information. However, to comply with the regulation, app stores make it more difficult for the developers to track end-user devices to get personally identifiable information. While this benefits consumer privacy, it may result into huge losses for the app developers. According to an interviewed developer organisation, a developer offering an ad-driven application can generate up to 50% more revenues from the end-user, if their device and IDs are known.

Speaking about data for the downstream business processes of app developers, they actively work with data on their performance, consumer behaviour data, business analytics, market trends analysis, comparison against competitors and so on. For the developers, one of the key purposes of this data is app store optimisation, meaning higher app store ranking. Most of it comes from the app stores. As illustrated by the platform case studies, both App Store and Google Play tend to provide very similar levels of access to such information for developers. The data and analytics related content and functionalities provided by the two largest app stores differs rather marginally:

- A/B testing functionality is built-in on Google Play Console, while developers on Apple App Store need to use external providers for this.
- More detailed data on user behaviour by country is provided by Google Play Store.
- More detailed data on purchase dates is provided by Apple App Store.
- Apple App Store claims to apply stricter control for data privacy.
- Play Store provides more metrics on comparison with competitors and wider app markets (in aggregated form).
- Google Play seems to provide more resources on tips and market insights to their developers than the Apple App store.

Furthermore, among other groups of platform business users, app developers can be characterised by greatest capacities (i.e., skills and technology) to run analytics on their business performance, make insights and develop tools helping to further develop their products, monetisation models and marketing strategies. This allows to substantially complement the data and basic analytics that app stores provide themselves. As an association representing app developers argued, many of them have developed their own “recipes” for app store optimisation, and this provides them with comparative advantage even against the bigger app companies with significantly more resources. In addition to this, many analytics providers offer services specifically to app developers, promising to help them better understand their users, user habits and behaviour (see more in the following section).

---


Although requests for wider access to the consumer/end-user data (i.e., users of other apps on the store) is increasing, an association representing smaller developers claim that these requests come primarily from larger tech companies. According to them, if data on end-users of other apps is not “handled carefully” by app stores, this would be detrimental to small tech businesses. Sharing data on end-users to all the developers active on the app stores would likely give advantage to bigger and richer companies. These could use this competitor data very effectively, harming the small developers who are currently used to navigating the app stores rankings with their own “recipes” for optimisation. The current situation of data access seems to be optimal for them, providing enough of both opportunities and protection.

5.2. App store optimisation tool providers

App store optimisation (ASO) is a process through which app developers optimise apps to rank higher in an app store’s search results and improve the click-through rate. It involves improving the visibility of an app within an app store. The ecosystems of companies that have clustered around the main app stores include a plethora of ASO tools, analytics and insight companies, servicing app developers from various app segments. They collect and sell data, analytics, and insights on business demographics, such as businesses’ and app products’ identifiers; and performance data, such as apps downloaded, global rankings and revenue. These are then used to produce app review and sentiment analysis, historical market, competitor, or performance trends. Some of these companies also collect data on consumer demographics and behaviour. In an aggregated form, these result into products such as app analytics, app optimisation, market research, app information, providing a good overview of the specific app markets and competition. Using this information, developers can optimise a mobile app’s marketing copy (keywords, title, icon, screenshots, description, etc.), functionalities, marketing, business strategies, monetisation models and so on.

Such third-party analytics and insights are important given that, just as in the other reviewed segments of platform economy, business users receive very little information on their competition from the platforms. Data providers for app developers, such as AppAnnie, Apptopia and SensorTower, fill this gap by presenting comprehensive app market data, including performance of specific competing apps and markets. Interviewed app developers mentioned that they use these sources together with the app store data extensively. The wide use of such services is corroborated by the use statistics provided by ASO tools themselves. For instance, App Annie has around a million registered users/apps, and it is just one of the key analytics and insight providers in this market.

In addition to this, just as in case of other sectors of platform economy, many app developers use the services of third-party insight providers to better understand their own customers and performance of monetisation strategies. For instance, companies such as Amplitude provide solutions based on the developers’ own data, aimed at profiling users to increase their engagement and optimise conversion. App Figures, which also provides ASO services, insights into market and competitors, offers data products to track the performance of individual apps of the app developer in one place.

Meanwhile, the app stores claim not to provide any data for reuse for businesses outside their company ecosystems. They recognise the automated data collection activities going on their app store pages, but do not actively support, nor restrict these in any way (although these are not allowed under the app store

---

47 https://www.businessofapps.com/marketplace/app-annie/#2
49 https://appfigures.com/
terms of use). In fact, most of the data that ASO providers get from platforms is collected through web scraping, gathering data from individual developers and through their own apps that collect data from consumers directly. Some ASO companies also list on the app stores their ancillary apps aimed at customer data collection, which app stores permit as long as they meet the app quality requirements (see the box below).

**BOX 2. APP ANNIE DATA COLLECTION**

App Annie, the leading market insight and ASO service providers to app developers, reportedly uses several different approaches to data collection.

First, to collect data on the audience of different apps (User Intelligence), App Annie distributes high-quality free apps to large numbers of end-users. These often are VPN software or mobile data management apps, which create a lot of value for the user, but also serves data collectors as a means to attract end-user opt-in. The software is free to download and use for the consumers, but only if they give the permission to collect data on the consumers’ apps, app usage, and other behaviour. Combining data from various such apps, App Annie gets massive reach.

Second, the company has developed data technologies that aggregate, anonymise and analyse the digital footprints, that individuals leave when they download apps, leave reviews, and comment on social media, at a massive scale in order to construct a clear demographic profile of a mobile app’s user base. This data is then supplemented by data from proprietary consumer panels and publicly reported data.

Third, App Annie gathers sales data from individual developers who use its services themselves. For example, once a developer connects their App Annie account with their iTunes Connect account, the sales data is automatically downloaded and processed. In exchange, app developers get free basic analytics. App Annie combines all the download and revenue data it has access to together to build a global model of sales.

Finally, that data is then correlated with the public ranking data in the top app lists to place sales figures, collected through web scraping. This involves breaking down the algorithms the app stores use for ranking. From there, App Annie can estimate downloads and usage.

Moreover, app store optimization services by third parties are often integrated with additional analytics services focusing on improving specific apps, or specialized app marketing agencies (like Preapps). These also collect app store data from the business users who are also using the services of these third-party providers.

---

50 For example, [https://www.mydatamanagerapp.com/](https://www.mydatamanagerapp.com/). As of April 2020, it is ranked first among mobile data management apps on the Google Play Store, with over 10 million users. Its publisher App Annie Basics presents itself as “the leading global provider of mobile market estimates”. Their apps help them “to create market research based on mobile behaviour”.


53 [https://www.preapps.com/](https://www.preapps.com/)
5.3. Other platforms

Most major digital online platforms also have their apps, listed on the app stores. These, therefore, can be treated as app developers (described above) in terms of the data they need and receive from the app stores. Besides this, some of the major platforms participate in various initiatives requiring data sharing between them. However, such initiatives involving Google and Apple seem to be ad hoc projects, rather than continuous collaboration.

For example, the ongoing Data Transfer Project (DTP) was launched in 2018 to create an open-source, service-to-service data portability platform so that all individuals (i.e. platform service end users) across the web could easily move their data between online service providers whenever they want. It was encouraged by the data portability requirement of the GDPR. The current contributors to the Data Transfer Project are Apple, Google, Facebook, Microsoft, and Twitter. The project is also open for other companies to join. The project aims to enhance the data portability ecosystem by reducing the infrastructure burden on service providers (i.e. platforms). The protocols and methodology of DTP will enable direct, service-to-service data transfer with streamlined engineering work.

The participants expect that this ultimately should stimulate competition and innovation. According to Apple’s representative, once the basic parameters of the Data Transfer Project are in place, and this “will not be an exclusive club”. It will rather set the standards, and subject to consent of the data subject, all participants will have access to data.

Another instance of ad hoc collaboration on data sharing between Apple and Google was started during the preparation of this analytical paper. In the context of the COVID-19 health crisis, the two tech companies announced “a joint effort to enable the use of Bluetooth technology to help governments and health agencies reduce the spread of the virus, with user privacy and security central to the design”55. Basically, Apple and Google will provide tools that will enable combined device tracing across both iOS and Android devices, which will mean that no matter what device a person is using, a central process will be able to track it, based on proximity signals via Bluetooth. Health authorities will then be able to develop apps that can trace which users have been in close contact with others, potentially infected with COVID-19. The process will be opt-in, and will require users to download a new, official app56.

54 https://datatransferproject.dev/
6. Data in the online travel services sector

According to the World Trade Organisation, tourism is one of the most dynamic economic sectors\(^{57}\). Statista estimates show that the numbers of travelers increase every year, and that over 1.6 billion people will travel in 2020\(^{58}\) (to compare, the figure was 674 million in 2000, and 1.23 billion in 2016\(^{59}\); although the figure is likely to be smaller given the travel restrictions imposed in response to the COVID-19 crisis). As of 2016, tourism industry generated USD 4 billion per day\(^{60}\).

Tourism-related services have numerous links to other major segments of the economy, as well as important societal and political implications. For example, tourism services are labour-intensive, and provide an important contribution to employment (World Tourism Organisation estimates that it creates 1 in 11 jobs globally\(^{61}\)). Recent developments show that some tourism services such as short-term accommodation rentals also raise serious concerns about the housing market in many European cities.

Digital platforms are gaining increasingly large role in the tourism sector, as more and more tourism services are purchased online. Statista forecasts that in this market, 70% of total revenue will be generated through online sales by 2023, compared to 66% in 2019. Due to their societal relevance, digital platforms in travel sector, compared to other sectors examined in this paper, are more visible and often more contentious (as illustrated, for example, by various movements against Airbnb globally\(^{62}\)).

Hotels and vacation rentals are among the key business users of online platform intermediation services in the travel sector. According to a survey by Hospitality Europe (Hotrec), in 2017, on average 26% of hotel bookings in Europe were performed via online platforms. The small hotels were significantly more dependent on OTAs than large hotel chains. Among the online travel agencies, the leader was Booking.com with over 67% of the market share in Europe, followed by Expedia Group with around 17% market share\(^{63}\). These shares, according to Hotrec’s representative, have since increased, although they could not provide precise numbers\(^{64}\). This significant market power puts the platforms in the position to unilaterally determine the scope and scale of data sharing, which is sometimes seen as too restrictive or even unfair by the business users.

---


\(^{60}\) World Trade Organisation (n.d.). Tourism and travel-related services. Available at https://www.wto.org/english/tratop_e/serv_e/tourism_e/tourism_e.htm


\(^{64}\) Their 2020 hotel survey was postponed due to the COVID-19 crisis.
These platforms - online travel agencies (OTAs) - neither directly provide the accommodation service, nor invest in physical infrastructure. Instead, their key stock-in-trade is data. It is likely that the major OTAs hold the most comprehensive datasets on travel sector and traveler behavior. These data are or could be used by different actors for different purposes; and they are capable of describing phenomena and experiences within real communities. OTA data are therefore are desired not only by tourism service providers - who contribute to the generation of this data - but also governments, NGOs, researchers and other actors.

A simplified chart illustrating the ecosystem of entities using OTA data is provided in the figure below. It builds on the case studies of five platforms in the travel technology industry: Booking.com, Expedia, Airbnb, Tripadvisor and eDreams ODIGEO. Although the platforms are characterized by slightly different business models and types of business users, the core principles of their data sharing seem to be similar. As the figure shows, business users have some possibilities to acquire consumer data themselves, but OTAs process much larger flows of customer personal and behavior information. Nonetheless, data generated by OTAs though intermediating transactions between accommodation/tourism service providers and consumers is widely used by a variety of entities outside the platforms. Further in this section we present each type of entities in more detail.

**FIGURE 4. ECOSYSTEM OF OTA DATA USERS**

![Diagram showing the ecosystem of OTA data users](image)

Source: developed by the authors. Note: this is a simplified visualization of platform data flows, revealed through desk research and the interview programme; it should not be viewed as exhaustive.

---


6.1. Hoteliers and short-term accommodation providers

The main users of data generated by OTAs are their current business users, especially accommodation providers. They need platform data mainly for the downstream business processes. Based on the interview and desk research, the business users of hospitality sector have relatively lower data needs, as compared to e-commerce sellers and especially app developers (although this varies notably by the size of enterprise). Nonetheless, the calls for more OTA data access in the industry are very prominent, and are based on the argument that the data is “co-generated” by hotels and OTAs working together.

Generally, the OTA data currently provided to their business users concern several distinct types of indicators, used for distinct purposes in tourism business management.

First, one of the main areas of data application in their business are dynamic pricing strategies. Efficient management of hotel revenues is related to supplying and selling the right rooms, to the right clients, at the right moment, for the right price. Otherwise these businesses face the issue of perishability, as an empty room from yesterday cannot be sold tomorrow and generate revenues. Given the increasingly dynamic market conditions, using flat pricing would be detrimental. Efficient revenue management, therefore, requires hoteliers to have access to continuously updated metrics, including financial information, market occupancy rates, market average rates, marketing costs per booking, OTA conversion rates and performance relative to customer segments. The major OTAs, including Airbnb, Booking.com and Expedia do provide their business users with analytics and suggestions for price optimisation, and argue that based on their evaluation these help business users to increase their revenues. However, these price analytics are viewed by some the business users as biased towards platform’s interests. For example, as illustrated in the Airbnb case study, some of the hosts notice that the prices suggested by the platform are too low; they would guarantee a booking on which the platform collects the commission fee, but this would mean lower profits for the hosts. This could at least partly explain the demand for third-party tools, such as AirDNA which also provides these services (see the section below).

Second, knowing one’s market and competition is key to attract customers and manage revenues. For example, to set competitive prices, hotel owners need to know how much other properties in that market charge. To define this market, hotels need information on location, average daily rates, accommodation types, quality/amenities offered, traveller types, and so on. Interestingly, what further distinguishes competition in the accommodation sector (and influences the data needs) from, e.g., e-commerce or software sectors, is the geographic scope. Although this market is very competitive, hotels compete for customers within a limited geographic area, usually within a well-known set of competitors (while online retailers, and especially app developers, often distribute their products globally). The data that hotels need on their competition or market trends, therefore, also primarily concern defined locations in which the properties are located. The main OTAs provide their business users with the possibility to compare their properties against a set of competitors. However, some platforms allow business users to choose these sets for business users, while others do make this choice themselves.

All the OTAs focusing on hotels as main business users claimed that the level of data they provide to their business users is sufficient or even superfluous to many groups of users. Given that the vast majority of

70 Revfine (n.d.), “10 Tips To Increase Your Hotel Bookings Through OTAs”, Revfine. Available at https://www.revfine.com/ota/
properties on their platforms are small businesses, many of them do not fully exploit the possibilities provided by OTA analytics. Meanwhile, for the large hotel chains analytics are not sufficient enough to be actionable.

Third, analysis of market trends (e.g., seasonal trends, trends related to external events, such as COVID-19, Brexit, etc.) is important to make informed and timely changes to the booking strategy. With resulting insights hotels and marketers stand a better chance of engaging with guests. Moreover, large hotels use this data to drive engagement, guest loyalty, direct booking, and ancillary revenues. All analysed OTAs seem to invest in studies (often outsourced to specialised organisations) or travel sector trends and market overviews, based on consumer surveys and behaviour data. These are often shared not only with their business users (with the goal to provide additional insights and tips), but the wider public as well.

Finally, the type of data that hotels seem to be most interested in the nominal customer data and contact information, allowing hotels to have a direct relationship with their customers prior to the visit. It is also the most contentious data-related topics in the discussions between OTAs and hoteliers. Both small and large hotel representatives argued that getting in touch with the customers is very important to provide tailored high-quality services (while using the OTA internal messenger tools, according to them, is too complicated, and the OTAs do not always pass the messages through). The only thing that hotels lack in order to provide a very personalised service, according to them, is knowing preferences of the individual customer. For example, many properties have the capacity to make arrangement for accessibility, adjust the lighting, pillows, meals, room arrangements or other aspects for the specific client, but they do not know the client’s needs. Providing or not providing personalised service, in turn, is related to hotel image and reputation, and OTAs do not provide enough opportunities to get the needed information. Therefore, hotel representatives claim that by not providing this information “OTAs hijack hotels’ relationships with the guests”. Moreover, nominal customer data would allow hotels to deal with fake or undesirable reservations. With the current OTA approaches, they cannot know if an undesirable guest has made a booking before they show up at the property.

In addition to these uses, large hotel chains see customer name and contact information as a means to build guest relationships with loyalty marketing. It is also crucial for targeting the right profiles with promotions, and rewarding the best guests with campaigns that encourage loyalty (both important aspects of their business models). Some real-world examples show that loyalty programmes are a very important tool for hotel chains to secure clientele and acquire new frequent customers. Access to customer contacts help to facilitate easy direct booking campaigns, which allow hotels to bypass online and offline travel agencies. The largest global hotel chains (such as Marriott, Choice, Hilton, IHG, Wyndham, Hyatt) estimate that direct-booking campaigns — even those offering discounted rates to consumers — are generating more revenue for hotel owners than bookings via the OTAs. Naturally, the direct hotel bookings undermine the business model of the online platforms.

All the reviewed OTAs except TripAdvisor and Airbnb do not share more than the minimum (needed to process the transaction) customer details with business users after the booking is made. Airbnb can be contrasted to other OTAs given that consumers on this platform have accounts that are displayed to their hosts. This provides hosts comparatively more consumer information, including some identification details, photos, as well as reviews about the consumers left by other hosts in the past (helping to identify fraudulent bookings). Meanwhile, Tripadvisor provides hotels that use its Instant Book service (see more details in the platform case study) with consumer identification and contact details that other OTAs conceal from their business users. What concerns the largest OTAs Booking.com and Expedia, large hotel

chains, are sometimes able to negotiate access to more customer data with these OTAs. This, however, is completely out of reach for the smaller properties. Booking.com and Expedia, try to compensate the lack of information about customer characteristics by providing anonymised consumer insight data on the guests that have booked on the property. Generally, however, most business users strongly complain about the lack of access to such customer data, and see this as one of the core issues in their P2B relationship.

Interestingly, multi-homing by accommodation providers is very common, and very often they have listings on several OTAs, in addition to their own websites. OTAs, in turn, acknowledge this market trend and provide APIs allowing hotels to manage their room availability, reservations, prices, and many other things within the hotel’s own system. On the one hand, multihoming is likely to shape the accommodation providers data needs: some of the key metrics provided by different platforms should be comparable to allow them to effectively manage their overall business. On the other hand, it can affect how different OTAs view data sharing and what metrics they provide (or, in case of nominal customer data - do not provide) in comparison to competing OTAs, in order to keep up in terms of the value proposition to hotels. For example, several reviewed platforms introduced their analytics dashboards for business users around similar time (2015-2016), and they constantly update them with quite similar metrics and functionalities. This provides some indication that they follow the developments in the competitor platforms.

Overall, while the interviewed OTAs tended to emphasize their effort in developing analytics for hotels in the interviews, the interviewed hotels tended to see them as a minor aspect of data sharing. It seems that for small businesses the OTA analytics are not that important, while large hotel chain have own capacities to use data to draw more comprehensive insights better tailored to their needs. Instead, business users in the travel sector of all sizes argued that they mostly need individual/ consumer level data and contact information.

6.2. OTA optimisation tool and travel insight providers

Data and insight providers specialising in tourism sector mostly focus on accommodation listings’ data. They use it to develop insights on business demographics, market trends, and business performance behind the analysed listings. Accommodation providers use these data services to access insights on pricing strategies, market performance, potential market locations, etc. Interestingly, however, while numerous optimisation tools exist for business users of platforms focused on short-term rentals (most notably Airbnb) and respective type of business users, almost none focus on providing such information to hotels, or on OTAs primarily listing hotels (e.g., Booking.com).

The interviewed OTAs claimed that they do not maintain any data sharing relationships with these companies. The optimisation tool or market insight providers, therefore, collect platform data by scraping their listing websites and other sources, such as information provided by hoteliers and hosts by linking their accounts, through consumer surveys, and so on. For example, Airbnb explicitly allows its hosts to transfer their account data to such data companies.

An example of such an optimisation provider is AirDNA. It is offering its clients - owners of short-term rentals listed on online platforms - a variety of data analytics services. The company uses listing data to offer metrics for optimisation of the rental business, aggregating data on every Airbnb rental worldwide, as well as listings of other OTAs (including Booking.com, Home Away). Its wide coverage helps accommodation providers to understand their own market and other market trends from different
AirDNA is also linked to another business, like Rentingyourplace.com, which offers consulting services to prospective Airbnb hosts. It is also not the only analytics company to mine platform data: others include, Beyond Pricing, SmartHost, Everbooked, and PriceLabs and many others.

Other tourism sector insight providers, such as Phocuswright, STR, Tourism Economics and Skift, focus on the tourism and accommodation sector more broadly. They also use scraped OTA data, as well as the public annual reports data to provide analytics and insights into the overall travel industry (online and offline, instead of specific platforms), and sell datasets and reports. As the interviews and websites of the mentioned providers indicate, their products are used by platforms and their business users alike. The information that they provide is important to run hotel business, but does not provide information related to doing business on specific online platforms. Nonetheless, given the widespread incidence of multi-homing in the market, such broader market overviews (rather than focus on individual platforms) are likely to be even more useful and relevant for the platform business users.

As these companies depend on accessing platform data and drawing valuable insights from it, their business model could be jeopardised should the OTAs close the access to such data collection. Nonetheless, several relevant instances in the past show that they have successfully overcome similar challenges. For example, in late 2015, Airbnb closed access to its real-time reservation data. AirDNA had used it to provide its services, and managed to train its algorithms based on this data. Since the access has been closed, the company uses the algorithms with the data scraped from the platform’s website and gathered from registered accommodation providers. AirDNA argues that their algorithm has an error margin of only 5% and therefore provides very precise information, which can be further used by accommodation providers to estimate the future demand and set appropriate prices. Nonetheless, the strong dependence of such companies on the OTA policies and practices remains. Their business model would be especially hurt if OTAs started to provide similar analytics or to sell access to their data (actual observed, not inferred/extrapolated as in case of many such third-party companies) themselves.

6.3. Governments, cities, NGOs, researchers, journalists

As described above, tourism and especially short-term accommodation - both central to OTA activities - are issues of high societal and policy relevance, especially due to their impact on the housing markets. Therefore, not-for-profit entities, including public authorities, activist groups, researchers and journalists are interested in and regularly use OTA data as well.

Previous research into OTAs confirmed that access to such data on platform companies can be difficult to obtain, as platforms themselves are reluctant to provide it. The impacts of short-term rental platforms on the availability and affordability of long-term accommodation in cities has led to numerous studies and reports:

---

74 beyondepricing.com
75 smarthost.co.uk
76 https://www.everbooked.com/
77 https://hello.pricelabs.co/
78 https://www.usewheelhouse.com/
79 https://www.phocuswright.com/
80 https://str.com/about
82 https://skift.com/
reports, many of which also point out the lack of easily accessible data. Browsing the OTA websites and scraping their front-end, therefore, is the key source of platform data to most actors (although in rare cases researchers follow a similar approach to the optimisation tool companies, and target platform business users to collect data). Scraped data has formed the basis for a number of studies and reports about short-term rental platform impacts. The journalists or researchers may scrape the data themselves or they use data scraped by others.

In 2020, a new data sharing agreement was announced between the major OTAs and the European Commission. Airbnb, Booking.com, Expedia Group, and Tripadvisor - four key competitors in the short-term rental market have agreed to share data about a global market with Eurostat. The first statistics should be released to public by the end of 2020. A representative of travel technology association explained that for platforms the main rationale to provide this data was to respond to the claims that short-term rentals are distorting the housing markets in the major European cities. In fact, OTAs (especially Airbnb) have recently faced a number of attempts to limit its activities by cities like Amsterdam, Barcelona, Berlin, Paris and others. OTAs also see value in Eurostat becoming a centralized body processing and publishing this data. Before this agreement, the platforms had been receiving lots of requests from cities or regions to access their policy-relevant data on short-term rentals and bookings - something all the OTAs viewed very critically in the interviews. The public bodies, according to the platform representatives, collected similar data from other OTAs, and did not process it properly (e.g., did not take into account the fact that multi-homing is very widespread in this sector, so the same hotel was counted as several), leading to exaggerated statistics on short-term rentals market. They expect that Eurostat will solve this issue.

6.4. Other platforms - the gatekeepers

Interviews with OTAs revealed advertising companies as another major player potentially re-using OTA data. This specifically concerns Google - the most powerful player in internet search, through which most of OTAs online traffic comes in. Not advertising on Google, as all the interviewed OTAs claimed, is simply

---

86 Ibid.
95 It is also important to note that in some EU countries, government authorities are also interested in hotel data for tax compliance enforcement purposes (e.g., there is an obligation for OTAs to share data in France; similar legislation has been introduced in Spain; this relates to exemptions of personal data sharing under the GDPR). However, OTAs aim claim this should not be their role, and it violates the rights of their business users. In fact, a recent litigation in Germany (VGH München, 20.5.2020, Case 12 B 19.1648) was won by Airbnb against the city of Munich, rejecting the city’s requests for hosts’ data.
not an option in their line of business. Although OTAs recognise that Google advertising products are very effective, they argue that the price they pay in return is too big.

Using Google AdSense and other advertising products, according to an interviewed OTA, means that they have to “open the door to their website”: the website’s content and ability to track individual visitors. The Google’s ability to draw behavioural insights from the data they get by tracking individual internet users on the specific websites allows it to develop a very effective advertising service. However, this also means that Google knows more about OTA visitors than OTAs themselves: for example, while OTAs can track only registered users on different devices, Google tracks everyone using Chrome, Android devices or other Google products on different devices and can link specific behaviour to individual consumers.

In addition to this, the service contracts between Google and at least some of the OTAs are prepared unilaterally and do not prevent Google from using data provided by or collected on the OTAs to re-use it for own purposes.

The OTAs have become especially alarmed when Google introduced its own local search functionality - a specialised vertical, competing directly with the OTAs in the ranking of Google Search results. When a consumer searches for a city or a hotel on Google (as most consumers do), the first results provided are ads, followed by Google’s “local box”, and only then by organic search results (see the figure below). After clicking on a property displayed in the local box, consumers are directed to another page, in which they see price comparison of that property from several OTAs. Then they can be directed to the specific selected accommodation provider on a selected OTA website.
As OTA representatives explained, in order to appear on that list with price comparison, OTAs must submit a lot of additional data on their room availability and prices to Google. It is feeds into Google’s advertising algorithms, aimed at tailoring and displaying the ads. However, as mentioned, their contracts with Google, do not preclude it from using this data for other purposes. OTAs, in turn, claim that Google does not provide them with important information on the users, traffic, conversion and insights on how to make their advertising bids more efficient.

Moreover, the Google’s local box limits the capacities of the OTAs to collect user data, as they are losing organic traffic on their websites. When consumers find the offer that they prefer on Google, they are directed to a specific listing on a specific OTA, instead of browsing the websites of selected OTAs at first. Therefore, all the user behaviour data remains with Google.
7. Analysis of incentives and constraints to data sharing

The analysis carried out for this study shows that platforms provide data to their business users, which is sufficient to process transactions, manage the business, and make some important business decisions. Nevertheless, some data desired by businesses is not provided, for a variety of reasons. As the platform case studies and sector overviews show, the current level of data access provided for platform business users can be viewed as shaped by finding balance between several interests that these platforms face.

— The interviews with app stores revealed that in the data provision to developers, the platforms aim to reconcile their goal to help the app developers succeed on the platform, and the goal to comply with data protection obligations and to meet user expectations related to the customer data use. Both are directly linked to the platform reputation and revenues.

— OTAs face similar trade-offs involving consumer privacy in what concerns behaviour data and analytics. Nonetheless, when it comes to nominal and contact data of consumers, OTAs balance not only the need to ensure successful stays at the hotels and consumer data protection, but also their objective to prevent hotels from bypassing their platforms to arrange bookings directly. The OTA business model would no longer be viable if this becomes widespread.\(^{93}\)

— In e-commerce, the same consumer privacy versus success of the business users/volume of transactions dilemma is common for most marketplaces. The aspect of consumer privacy as a constraint of data sharing, according to the international platforms, is further strengthened by legal uncertainty, and diverging interpretations and enforcement of data protection regulations in the countries in which they operate. In addition to this, even though the evidence from case studies is not sufficient to claim that vertically-integrated platforms systematically tend to provide less data to its business users, there is likely a difference in the incentives to share data of vertically integrated and non-integrated marketplaces. The former may also take the success of their vertical into the equation.

In addition to these, issues of technical implementation, investment in data processing and others concern most if not all the overviewed platforms. Meanwhile, the general role of data in platform business models and competition becomes important regarding the incentives and constraints for platforms to share their data with other businesses, including other platforms.

However, ultimately the distinction with whom data is shared - be it business users, other businesses or other platforms - is less important. For example, if a platform opens a data flow primarily intended to their business users, it cannot prevent other actors from getting access to this data as well (as illustrated by the activities of optimisation companies who collect business user data). Therefore, the discussion of incentives and constraints for data sharing – legal, economic and technical - is presented from the perspective of general openness of platform data. This chapter further investigates it in more detail.

\(^{93}\) Associations of both large and SME hotels, in turn, argue that such arguments are not substantiated. According to them, no proof exists that this would not be damaging to the OTAs, and consumers would still use them when panning trips to new locations. Meanwhile, accommodation providers should have the direct consumer relationship as they are the ones to provide the service.
7.1. Economic barriers and incentives

Interviewed online platforms and associations representing them listed several economic factors, determining the current levels of data openness to other businesses. They all can be also explained from economic theory and economic rationality point of view, according to which profit maximisation is often seen as the key goal of companies. In this view, incentives and constraints of platform data sharing largely depend on their business model. Three aspects of platform business models are important:

— First, they are inherently built around data.
— Second, the platforms overviewed in this paper generate revenues through facilitating transactions between their business users and consumers (i.e., transaction-based fees).
— Third, some intermediating platforms are also vertically integrated, meaning that they compete on the same marketplace as their business users.

We further overview what incentives and constraints to data openness they introduce.

**Data as the core of the business model.** Generally, the limitations to which companies are willing to grant access to their data (and specific types of it) to third parties mostly depend on the role of data in their business model: whether they use it to improve the business, to develop new verticals, to monetise through behavioural advertising, or other. If the exploitation of a particular dataset is at the core of its business model, an online platform will be unlikely to share it\(^\text{94}\). In many aspects, this is indeed the case. Data is the backbone of any tech business, which can explain the general reluctance of platforms to share their data\(^\text{95}\).

Platforms have interest in keeping their systems closed due to the dependence on their user base. In the meantime, competitors and new entrants want access to user data so that they can provide competing and complementary services\(^\text{96}\). Data openness, therefore, is risky. For example, one European platform mentioned that they detect abusive bots originating from China\(^\text{97}\), crawling their pages or trying to use their APIs. At the same time, they see Chinese marketplaces as serious competitors, disregarding all the regulations European companies must comply with, and therefore gaining competitive advantages. Furthermore, intense competition between European and/or American platforms exists as well.

Moreover, it is costly to collect, produce, observe or infer the data. Platforms have made significant investments into the development of their data collection and processing capacities. Opening it to other actors could allow them to freeride on the investments that platforms have made.

**Platform as an intermediator.** Some data sharing with third party businesses - within certain limitations - contribute to the goal of profit maximisation. The main feature of all the platforms overviewed in this paper is that the most significant share of their revenues comes from the commission fees charged on transactions between their business users and consumers. Therefore, to succeed economically, platforms need to be attractive to both sides of the market: business users and consumers. Only platforms with a good value proposition for both types of users can scale and exploit direct and indirect network effects.

To be attractive to **business users**, platforms strive to offer a useful and attractive set of data and analytics. This seems to be an important factor in the three analysed sectors of the platform economy: e-commerce,
hospitality and app development. For example, the interviewed OTAs said that metrics and dashboards provided for business users is part of their service package which is continuously reviewed to remain attractive among competing OTAs. Various blog posts for e-commerce sellers, listing pros and cons of different e-commerce marketplaces, also often use data and analytics provided as one of the judgement criteria. Generally, platform business users often want more data, so if a competing platform invests in analytics, others feel the pressure to do so too, in order to maintain their business user base. Several interviewed platforms emphasised this among the key incentives to invest in analytics capacities for business user-oriented analytics and insights.

Moreover, economic success of business users on the platform - in terms of the volume of transactions made - automatically means more revenues for the platform. High quality analytics and insights, if used properly, can help business users to develop better products, services or business processes, allowing them to succeed on the platform and therefore to generate more revenues for them. This factor was also mentioned by platforms from all the overviewed sectors as one of the incentives to invest into analytics for their business users. It may also explain their tolerance to optimisation and insight companies scraping platform data or gathering it from other sources: ultimately, these companies help the business users to make more sales on the platforms. At the same time, however, there is an incentive that the type, granularity and presentation of data provided by platforms would help businesses succeed on that particular platform rather than other platforms. This potentially significantly limits the characteristics of what data is shared.

Consumer interests, especially what concerns their data privacy, are another part of this data sharing equation. All the interviewed platforms emphasised customer satisfaction among the key goals of their data policies. Generally, consumers want their privacy protected, and all platforms strive to ensure that personally identifiable consumer data is not accessed by any third parties. For some of them, like Apple, exceptional data protection and privacy approaches are part of the brand’s DNA. It is therefore an important factor limiting the extend of platform data sharing to their business users.

E-commerce and accommodation platforms also expressed a concern that if the sellers or accommodation providers get customer email information, they will flood the customers with marketing messages. They also are sceptical about the idea to introduce the business users with a possibility to obtain customer consent to process their data themselves (although this is strongly advocated by platform business users in tourism sector), as this main introduce additional burden on consumers or might confuse them.

Furthermore, the raw data generated by platforms is problematic because of its volume, personally identifiable nature and usability. Combining the two types of incentives discussed above - supporting business users and satisfying consumers - results into the necessity to properly anonymise and process the data. In fact, unprocessed data would not be very useful for the majority of business users, as they lack analytical capacities to make use of it. Many interviewed business users - with the exception of large companies - who reported feeling disadvantaged by limited data access, could not provide a detailed vision of how they would analyse the raw data from platforms and draw insights to help their business. Opening data, therefore, would likely be more advantageous for large companies, which would get additional means to outcompete the small ones.

Providing processed data, in turn, requires notable investments for platforms into data collection, management, analysis and presentation. Platforms are incentivised to invest only as much as this investment could ultimately translate into more revenues. This, however, is impeded by the fact that some

---

98 For example, McGregor, L., “Amazon vs eBay: which marketplace should you sell on?”, Edesk. Available at https://blog.edesk.com/resources/amazon-vs-ebay/

99 This concern was expressed by representatives of small tech companies.
businesses, such as small family shops on e-commerce platforms or small hotels on OTAs do not use this data at all, even if provided in the form of easy-to-understand insights.

Another limit to the extent of data provided by platforms to their business users concerns the possibility to bypass them as intermediators in transactions with the consumers. For example, OTAs argue that if they provide customer contact data to establishments that are listed on these platforms, the establishments will be able to bypass them, undermining their business model. Similarly, e-commerce platforms do not share customer identification details to avoid businesses targeting the customers through ads and directing them to their personal websites, bypassing the platforms.

**Vertically integrated platforms.** Finally, from the economic perspective, vertically integrated platforms may have another reason to avoid data openness. They would not share data that would allow their business users to become competitors. Depending on other variables, such as the share of revenues generated by the marketplace (as opposed to the company’s vertical segment competing with the business users) investment and competition, these limits of data provision could be expanded or lowered.

### 7.2. Legal barriers

Generally, data protection and competition legislation are factors strongly shaping the current practice of platform data sharing with other businesses. This was voiced by the interviewed platforms, as well as industry associations representing both platforms and their business users.

First, when business users request more data from platforms, they effectively request access to data about other businesses active on the same platform. This has repercussions from the perspective of competition policy. For example, sharing seller-specific revenue information among sellers can be interpreted as providing a competitive advice under the German anti-trust law.

Second, most of data that platforms collect is personal in nature. Long before the GDPR, the key guiding principle in handling personal data had been that as few entities as possible should be involved in data processing. The fact that platforms, collecting loads of personal data are not sharing it with other parties is in line with this regulation. Some of the platforms examined in this study, especially the app stores, also play an important gatekeeping role in protecting consumers from abuse of their data by app developers, some of which may wish to use it for unethical or even illegal purposes (as discussed in Chapter 5).

In fact, the interviewed platforms expressed that their key concern related to data sharing with business users is meeting the requirements of the data protection regulation. Some platforms indicated that they are willing to share more data, but legal uncertainty and the data protection regulation are among the main obstacles for this. For instance, a variety of data protection regimes globally, as well as diverging interpretations and enforcement of GDPR across Europe have been reported as a substantial constraint by platform interviewees from multinational platforms. In addition to this, while in the public discussions on data sharing the main distinction is between personal and non-personal data processing, from the business perspective, this distinction is hard to make. On the one hand, it is because of the broad definition of personal data in the GDPR. On the other hand, because at operational level companies combine various types of data. Therefore, GDPR possibly limits not only personal data sharing but data sharing in general: data is anonymised or pseudonymised after combining it from various sources, and there is an inherent risk that non-personal data can be re-personalised.

From the business user perspective, several interviewed business associations (representing tech and hospitality sectors) argued that GDPR has resulted into even more power to the large online platforms, who now use it as an argument to restrict access to the vast datasets of behavioural and other consumer data. Interviewees from tech and hospitality industries suggested that GDPR should be revised to allow
platform business users access more customer personal data held by platforms to be able to conduct some of their business activities (as explained in more detail in Section 6.1). Brands, app developers and hotels would also like to have more data on consumers, to be able to match this with some data from data brokers and develop more effective targeting and advertising. However, this could well be detrimental to consumers. Existing evidence shows that although most consumers only engage with personalised marketing (and this is a reason for businesses to require more personal data), at the same time they are largely concerned about their privacy. Therefore, the broader sharing of personal data among businesses could also be an undesirable outcome.

### 7.3. Technical barriers and incentives

One of the general factors for platform reluctance to share data concerns interoperability, and the lack thereof. This especially concerns data sharing between companies in the tech sector, including platforms themselves. The problem is also widely recognised in the EU policy discussions. In the Digital Agenda for Europe, the Commission has identified a lack of interoperability as one of the significant obstacles to a thriving digital economy.

Although there is no universal definition of interoperability, it can be understood as the ability of a system, product or service to communicate and function with other (technically different) systems, products or services. In the digital economy, interoperability issues typically relate to information exchange and data. In the context of platform data access, therefore, interoperability can be understood as the ability to transfer and render useful data and other information from platform to platform, or from platform to business. It is also a crucial precondition for data portability, mandated by the GDPR - that is the data subject’s ability to move data among different applications, platforms or online services.

At its very core, interoperability is a feature of technological design of systems (or platforms in this case). For it to work, platforms and other businesses have to work together and agree on common standards at the design phase with the intention of future data sharing. However, at the time when the structures and processes of the overviewed online platforms were developed, they did not consider interoperability with other platforms and businesses as a necessity. Generally, engineers write a programme for what it needs for its own business. Such reasoning results in code tailored to the requirements of one particular business. As a result, it is often unfeasible for a platform to work with its competitors, and even if the platforms do collaborate, they would do so under unilaterally.

Introducing interoperability between functioning systems is costly, because it requires revising the back-end code in order to adapt to common standards. Moreover, the participating platforms and businesses - often direct competitors - must collaborate to develop and implement common standards. Interoperability implies interrelationships between several systems, so these functionalities cannot be developed and operate unilaterally. Even once the needed technology is in place, participating platforms cannot be sure when their investments would pay off. Additional questions arise, such as the legality of data sharing, as regulation differs from country to country (for example, while personal data protection is generally strict in Europe, Chinese government has access to everything that Chinese platforms collect). Moreover, developing additional structures for interoperability in the systems make the system itself “heavier”, slower, and more expensive to develop further and make individual changes. This might limit possibilities for innovation within individual platforms.

In addition to this, greater degree of homogeneity (i.e. extent to which uniform standards and interfaces are used) limits the possibilities of

---


firms to develop their own specific products and services, because they must comply with these standards and interoperability requirements."
8. Possible solutions

As the analysis presented in previous chapters shows, platform data can be used by various entities and serve various purposes. Within the past decade, higher data storage capacities, faster means of data transmission and more powerful computing performance have significantly advanced the possibilities of data-driven innovation even outside the tech sector. Innovation, relying on the availability of huge amounts of training data and vast datasets, is in turn an important precondition for successful competition in dynamic online markets and beyond. These developments have intensified the calls for encouraging business-to-business (B2B and therefore P2B, P2P) data sharing in both recent legal scholarship and policy discussions.

The analysis showed that the current levels of platform data access is sufficient to process transactions, manage business process, and make more insightful business decisions. Some of the overviewed platforms, especially the app stores, also play an important role in protecting consumer privacy. However, at the same time platforms have enough exclusive data, which could have important consequences for competition and innovation.

Further we overview the solutions proposed in literature and by interviewees. We focus the analysis on the economic and technical barriers for data sharing identified, with potential repercussions for innovation and competition. Some of the proposals discussed seem rather complementary (rather than policy alternatives), and a combination of the could be considered for the new platform data openness regime.

8.1. Overcoming economic barriers

As explained in Section 7.1., economic barriers and incentives strongly shape the current platform behaviour concerning data sharing with third-party businesses. Companies in the tech sector - including platforms - have built their businesses around data. Platforms are reluctant to share data because it is at the core of their business models, and may be used by other actors to undermine their competitive position.

Data sharing between businesses (including both platforms and other companies) could be viewed from the game theory perspective as a prisoner’s dilemma. To illustrate (see the table below), the basic premise of the setup is that the direct cost and risks of sharing data or insights with another company amount to, let’s say -2 (e.g. considering the costs of implementation and revealing business-sensitive data), while having a partner that cooperates with you in this way imparts larger gains, represented by our number ‘5’ (e.g. considering access to bigger datasets and therefore more possibilities for innovation and business improvement). This cooperation can be the other businesses or platforms sharing their data within certain environment.

---

105 The framework has been applied to data sharing in other contexts as well, for example: Moore, A. D., & Martin, S. (2018). Privacy, Transparency, and the Prisoner’s Dilemma. Available at SSRN 3212217.
TABLE 2. PRISONNER’S DILEMMA

<table>
<thead>
<tr>
<th>BUSINESS 2</th>
<th>Cooperate</th>
<th>Don’t cooperate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BUSINESS 1</strong></td>
<td>Cooperate</td>
<td>(-2; +5)</td>
</tr>
<tr>
<td>Cooperate</td>
<td>(+3; +3)</td>
<td></td>
</tr>
<tr>
<td>Don’t cooperate</td>
<td>(+5; -2)</td>
<td>(0,0)</td>
</tr>
</tbody>
</table>

If the other business or a platform cooperates, both pay the price for sharing data or insights (-2), but gain the gains as well (5), amounting to a scoreboard of +3/+3 for each company. However, if one party cooperates (provides data), but the other does not reciprocate by sharing their own data or insights, the cooperating platform or business pay the costs but do not obtain any gains (-2/5), while the other company has just the gains of data shared with them and none of the costs. From a rational perspective, choosing not to cooperate at all (0/0) seems to be the most logical choice 106. This could be changed only if the platforms and other businesses trust each other to collaborate and expect gains that match or exceed these costs.

To encourage data sharing, therefore, platforms have to break through the stalemate created by the prisoner’s dilemma. In the context of P2B or P2P data sharing, few options may apply:

- Establishing regulation which would introduce costs for refusing to cooperate, therefore making this option less rational than cooperation.
- Reducing risks and costs related to data sharing, making the collaboration more attractive for both parties by guaranteeing payoff. Safe ways of sharing are part of the solution towards ensuring more data sharing, collaboration and reciprocity (which is especially important in this context; see the box below for an example). Part of this solution could be third-party intermediaries or technical solutions, ensuring the interests of both collaborating parties and establishing trust.

BOX 1. THE IMPORTANCE OF RECIPROCITY PRINCIPLE IN DATA SHARING REGULATION: AN EXEMPLARY CASE OF PSD2 DIRECTIVE

In relation to possible policy options to facilitate increased B2B data sharing, one stakeholder interviewee stressed the importance of ensuring the principle of reciprocity. This topic is being widely discussed in the context of the open banking initiative, and the revised Directive on Payment Services (PSD2) 107. The stated purpose of the rule XS2A 108 of the Directive was to allow the Fintech start-ups to access the users’ payment accounts data held by incumbent banks, in order to help them innovate and grow. However, PSD2 has been criticised for being an asymmetric data sharing framework, which provides access to data to the big tech players, whose business models are built on the power to extract value from data, while maintaining exclusivity over their...


108 Article 36 of the EU Payment Services Directive 2 No. 2015/2366.
own datasets\textsuperscript{93}. This is also what the growing evidence suggests: big tech may use the data to develop new services and enter new verticals, putting banks at competitive disadvantage. This, in turn, can potentially harm and reduce consumer welfare\textsuperscript{92}.

The banking sector has been vocal\textsuperscript{92} about the lack of reciprocity between banks and third parties, especially the big tech, and the associated risk of increased concentration of the market for data. They advocate for cross-sectoral regulatory approach to level the playing field. Among the proposed solutions is the introduction of "reciprocity clause", under which in exchange for customers’ payment information, big tech companies would have to provide banks with the behavioural data of the same customers (upon their consent)\textsuperscript{93}. According to the same authors, reciprocity would enhance data sharing, foster competition and create incentives for innovation among the market players. In Australia, the concept of reciprocity has been introduced in the Open Banking review, supporting reciprocal obligations that an accredited data recipient should be obliged to provide "equivalent data"\textsuperscript{94}.

We further review possible ways how to implement these options and what arguments for and against may apply.

8.1.1. Binding regulation and mandated data sharing

Currently, competition law is the only applicable general statutory law that can be applied in cases when lack of data access undermines market inter-industry competition. Nonetheless, it remains questionable whether competition law is sufficiently flexible to provide solutions for all platform data sharing related problems.

EU competition law prohibits businesses that are dominant in a market from abusing that position of dominance. An abuse of the dominant position can occur when a dominant company refuses to supply or provide access to essential facilities. 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 law.\textsuperscript{94}

Any regulation on data sharing beyond competition law must consider the incentives and disincentives of various players for data sharing. Looking through the game theory framework, compulsory data opening would increase the costs of non-cooperation. However, it would not make cooperation more desirable in absolute terms (as the benefits and other risks of data sharing would not change). A number of experts suggest regulatory measures requiring companies to make some of their data available to rivals


\textsuperscript{98} Graef, I. (2019). Rethinking the essential facilities doctrine for the EU digital economy. RJT ns, 53, 33.
(see the box below)\textsuperscript{115}. However, representatives of the industry as well as several researchers argue that currently no strong case for binding regulation for data sharing exists\textsuperscript{116}. According to the interviewed industry representatives, mandated data sharing could impede the development of new business models in such a dynamic environment. Further, mandatory data sharing may have a negative impact on platforms’ business and profits, incentivising 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.\textsuperscript{117} Moreover, according to Prüfer\textsuperscript{118}, even if mandatory data sharing was fully implemented in the EU, it is only a necessary, not a sufficient condition for lively competition and innovation on data-driven markets. Competing successfully requires more than access to the raw data those firms collect by logging interactions with users. Therefore, any new regulation should carefully consider what data should be opened by what platforms, to whom and under what conditions (see the box below).

BOX 3. PROPOSAL FOR MANDATORY SHARING OF USER INFORMATION IN DATA-DRIVEN MARKETS BY PRÜFER (2020)

In a recent paper “Competition Policy and Data Sharing on Data-driven Markets\textsuperscript{119}”, Jens Prüfer details a proposal to design a regulation mandating data sharing in data-driven markets. The author proposes that a number of questions must be answered in designing such a regulation, and proposes several pointers.

1. How to identify a data-driven market empirically?

Mandatory data sharing is unambiguously positive only if data-driven indirect network effects are sufficiently pronounced (i.e. if user information decreases the marginal cost of innovation substantially). To identify such data-driven markets, the author proposes first to develop a market definition. The market delineation is cumbersome, and many data-driven markets tend to overlap (e.g., that Google, Facebook, Microsoft and Apple are competing in the online advertisement market, as well as app markets). Therefore, a reasonable starting point for market definition is to take a user/consumer perspective and to ask what service the user demands, and which providers may have an offer that can satisfy the user’s needs. then, the demand side of the market should be studies (e.g. what drives users’ consumption utility), as well as the supply side (e.g. what drives measure of product quality).

2. Precisely what information should be shared on which market?

According to the author, only the sharing of user information is relevant, i.e., raw data about users’ choices or characteristics, which can be automatically (and hence at virtually zero marginal cost) logged during a user’s interaction with a service provider. The policy proposal explicitly does not include processed data, in which the original service provider/dominant firm already invested effort, for instance for data analytics (and hence at positive marginal cost). If such data would be required to be shared, it might facilitate free riding of smaller competitors and crowd out the dominant firm’s incentives to invest into analytics in the first place. Also, in order to disseminate the effects of data sharing as quickly as possible, user information would have to be shared on a «continuous» basis, that is, as frequently as technically possible.

\textsuperscript{119} Available at http://library.fes.de/pdf-files/fes/15999.pdf
3. **How can user information be anonymized and how can re-identification of individuals (technically or legally) be avoided?**

In order to mitigate the risk of indirect re-identification a data-sharing law could be introduced which prohibits the re-identification of individuals from shared data. Re-identification would establish an illegal act and would be punished. Further, privacy and data security concerns could be mitigated by building data protection into the data sharing governance structure. Instead of granting organizations direct access to shared data, a trustworthy data intermediary (data trustee; see more detail on data trusts in the following sections) could be established to be responsible for guaranteeing the compliance of data sharing with data protection rules. The intermediary could safely pool shared user information and ensure the fundamental privacy rights of the users by anonymizing the data before sharing it with eligible third parties.

4. **Who exactly should share data?**

The theoretical proposals that all firms in a data-driven market should share their user information with others seem suboptimal for the author. This is because (a) data sharing comes at a cost and creates an administrative burden; (b) large, dominant firms are more likely to have access to other sources of information that complement user information from this market and hence have higher marginal benefits from new user information received. As the goal of the policy proposal is to establish, as much as possible, a contestable level playing-field, the author suggests that large firms should share more data than small firms. He proposes that a threshold should be introduced, for example, such that the largest two or three firms have to share data, as these are the contenders for «dominant firm» status.

5. **Who should have the right to get access to the shared data? At what price?**

The author proposes that user information should be shared with every organization that is active in the respective industry or that can explain how it would serve users with the data. He suggests to draw inspiration from the Payment Services Directive 2 (PSD2) in the financial industry, which entitles third parties, with the consent of the account holder, to access payment accounts in order to initiate payment transactions via an internet application or to consolidate account information from one or more accounts into one application.

6. **How should data sharing be organized? What is the optimal governance structure?**

According to Prüfer, decentralized data sharing (e.g. via APIs) seems impractical. An intermediary is needed between senders and receivers (implementing centralized enforcement). This intermediary organization would be tasked with the structure and operation of the data-sharing scheme. It would have to cooperate closely with the data-sending firms in an industry, to validate the business plans of would-be receivers and to make sure that all certified receivers receive the appropriately anonymised user information of senders in a standardized, equitable and workable way.

Please see the following Section and the box on data trusts for other authors’ accounts on how similar bodies could function.

---

8.1.2. **Self-regulation: access to data under FRAND**

To incentivise sharing, a structure of relatively flexible rules aiming more at equal treatment of all involved stakeholders would likely be more appropriate than mandatory data openness, according to
some authors. When companies develop best practices and codes of conduct for data sharing themselves, this can incentivize sharing by lowering transaction costs and building trust\textsuperscript{120}.

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.

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\textsuperscript{121}. 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.

As the analysis presented above shows, the constraints of platform data sharing in some respects is similar to those that FRAND aims to solve in the context of SEP. First, concentration of data in the hands of several large platforms can potentially lead to the abuse of market power. Data itself in some respects is similar to SEPs, as both 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\textsuperscript{122}.

However, important differences between SEPs and data exist, which have important implications for the application of FRAND in this area. The main one is the nature of their respective legal protection. SEPs, just as other patents, are exclusive rights which create legal monopolies. Meanwhile, as of today, there are no ownership rights in data. According to the current research and policy discussion, exclusive rights may be even detrimental to the development of the data-driven economy\textsuperscript{123}. Instead, data can be considered exclusive if they are under the sole de facto control of a company, due to technological protection or access limitation. In other words, the platform which stores the data in its servers, "owns" the data\textsuperscript{124}, and has the twofold power to use it and prevent others from accessing it.

Another distinguishing feature of SEP is that the standard in which the SEP is included is based on an agreement between several market actors (i.e. members of a standard-setting organisation). This makes alternative standards in the market less likely. With regard to data, a situation similar to SEPs would occur if market actors could only rely on single-source data that is accessible only as a consequence of coordination between market actors that hold the data\textsuperscript{125}. While it is the case regarding many types of


\textsuperscript{121} European Commission, Towards a common European data space (Communication), COM(2018) 232 final.


\textsuperscript{125} Ibid.
platform data, the analysis also showed that for some types of data various sources and substitutes do exist (like estimations provided by third-party data companies, or open source alternatives for proprietary datasets, etc.).

In addition to this, as explained by legal experts in this field, a patent is disclosed and the buyers accessing it under FRAND “know what they pay for”. It is not possible to create the same conditions when it comes to data: one cannot be sure if the data will be useful before actually accessing and processing it. Therefore, FRAND is much easier to implement with SEPs, as it can more easily enable win-win situations, in which parties are incentivized to contribute to the common patent pool, while FRAND access strengthens trust and cooperation.

Moreover, according to ACT/ The App Association, there are more factors pointing to the fact that the dynamics and circumstances of platform data sharing are entirely different compared to issues related to intellectual property and SEPs. What concerns the application of FRAND, there must be a standard126, a voluntary agreement to license under FRAND terms as part of that standard, and an exclusive right (through the patent) by the FRAND licensor without competition. **This is not the case with platform data.** Furthermore, unlike in FRAND licensing of SEPs, there is a tension between sharing data with different parties and the privacy of individuals on which this data is collected.

Therefore, based on the analysis done by Richter and Slowinski127, FRAND principles are **not equally applicable to different modes/regimes** of data sharing for remuneration between platforms and other businesses. The authors consider three data sharing regimes with regard to the applicability of FRAND: direct data exchange, data pools and data sharing platforms.

First, in the case of **direct data exchange** between two companies, there seems to be little or no space for FRAND-based concepts. It is part of the data holder’s individual freedom to decide whether to share data and under what circumstances. In fact, our analysis shows that platforms usually share data **directly** with their business users and other actors in their ecosystems.

Situation is different regarding what the authors called **data pools**. 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)128. Companies can use data pooling to share data “in reference to a given service or generally in an industry, or within an e-ecosystem”129. As in the case with SEP to which the principles of FRAND are applied, companies share a commodity (patented technology or data) to their common economic benefit. Likewise, 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 under FRAND. However, the data itself in

---

126 However, the FRAND concept does not necessarily need to be limited to standardisation. In fact, FRAND is an increasingly accepted ‘good faith’ notion recognised by competition authorities and applied as a simple access remedy for the supply of a particular product. For instance, in the context of the Microsoft case the European Commission determined that Microsoft’s operating system APIs was an essential input and required FRAND-like access. [https://ec.europa.eu/competition/information/digitisation_2018/contributions/4ip_council.pdf](https://ec.europa.eu/competition/information/digitisation_2018/contributions/4ip_council.pdf)


this case should be considered critical for a specific purpose (e.g., data generated by a transport provider in a smart city context).\(^\text{130}\)

Moreover, in considering the application of FRAND to 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. The FRAND 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.\(^\text{131}\)

**BOX 4. DATA TRUSTS**

Data trusts are seen as a possible way to address the challenges of the data economy, including protection of consumers’ interests with regard to their data, issues that many organisations face in getting access to data effectively controlled by a few large players. The model, inspired by legal trusts, is expected to enable and enhance safe and legal B2B data sharing by developing repeatable terms for sharing data under mutually beneficial agreements.

While there is no common definition or consent of a data trust, the Open Data Institute (ODI) has defined it as “a legal structure that provides independent stewardship of data”\(^\text{132}\). The trustees, which have legally binding responsibilities and liabilities, are entrusted with making decisions about what data to share and with whom. From a legal perspective, a data trust is a mechanism for achieving a defined set of aims, including:

- to enable data to be shared;
- for the benefit of those sharing the data, and possibly also for some, broadly conceived, public benefit purpose;
- respecting the interests of those holding the data;
- ensuring the data is used ethically and in accordance with the rules established by the data trust; and
- ensuring that whoever holds data which is subject to the data trust rules does so safely and securely, and that data is dealt with appropriately (for example by deletion) if the data trust comes to an end.\(^\text{133}\)

Specific data trusts may also have one or more of the following aims and characteristics:

- collective management of individual rights and interests (including sharing of benefits received by the data trust);
- standard set of rules etc. to govern all data sharing and use (this can be subject to FRAND principles);
- custodian/steward making decisions on behalf of data providers/ data users; and
- ability to evolve to take on new purposes, governance and working methods.

The use of data trusts is considered appropriate by the ODI, if the group involved in sharing data is large, has diverse interests and may change constantly. Otherwise, data sharing between members of a small group of organisations can

---


easily be managed through bilateral or multilateral contracts, and the mechanism of a multi-party data sharing agreement is simpler and cheaper than setting up a data trust.\(^{134}\)

Data trusts are also viewed as measures to protect the interests of consumers by restricting data uses that are not in consumers’ interests (more so than data marketplaces), but also to enable new data flows that are currently held back by companies.

Although few data trusts or similar mechanisms exist in practice, the idea has been receiving increasing attention. For example, three pilot data trusts have been established by the ODI in the UK. The pilots examined whether data trusts could increase access to data while retaining trust and concluded that data trusts could provide a good approach to data-sharing where there are conflicting interests. Other countries, such as Canada\(^{135}\) and Germany\(^{136}\), are also considering the establishment of similar mechanisms.

The third mode of data sharing, as outlined by Richter and Slowinski\(^{137}\), is a data sharing platform, 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\(^{138}\)) 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. In such cases, the data sharing platform seems to be assuming a role similar to the standard setting organisation for SEP, while those companies who share data through the platform are similar to the members of a standard setting organisation or holders of SEPs. In such a case there is scope to exploring the use of FRAND principles.

The key issues to be resolved in order to operationalise FRAND principles for platform economy are the price of data, what ‘fair’ and ‘reasonable’ access to data is;\(^{139}\) and the role (if any) that the public authorities should take in facilitating cooperation and/or ensuring compliance.

8.2. Limitations on data use imposed on vertically integrated platforms

The interview programme and analysis of other evidence showed that the key issue related to risks to competition stemming from data advantage concerns vertically integrated platforms. Business users in e-commerce, as well as OTAs (as business users of Google) expressed concerns that vertically integrated platforms...
platforms such as Amazon or Google are in the position to use the data from their marketplaces and business users to develop verticals or improve and market their own products.

One possible solution aimed at preventing platform from gaining unfair advantage from these data is a regulatory action limiting how the data collected by the platform on their business users and consumers can be used. An example of this is “firewalls” on vertically integrated platforms such that the retail branch of a vertically integrated platform has the same level of data access as business users. Apple App Store and Google Play claim that this is already in place, and their first-party app developers do not have any preferential access to the app store data. In contract, whistle blowers report that Amazon Retail use the exclusive access to Marketplace data to develop private label products (see the case studies in Annex 1). In practice, given that identification of such platform behaviour and enforcement of such rules would be very complicated (e.g. it is hard to know how the platforms are using the data that they have access to), the reversal of burden of proof might be considered.

Reversed burden of proof means that a person (natural or legal) must be able to prove that they are innocent. For example, as required by the GDPR, rather than the data subject having to demonstrate justified grounds for objecting, the controller must demonstrate “compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject”.

In the context of digital platform economy, DG Competition is considering the reversal of the burden of proof in adjusting EU rules for large internet companies in competition cases. Digital platforms suspected of anti-competitive behaviour may be required, in certain cases, to demonstrate that their conduct is beneficial for their users (i.e., to establish the “pro-competitiveness” of certain conduct) to avoid legal repercussions. The change in the burden of proof would be limited in its scope and would apply to companies preventing users from accessing multiple apps and companies restricting access to their data for third-party applications looking to offer supplementary services — such as apps that read emails. A number of EU countries, including France and Germany, are also considering application of this principle in the context of anticompetitive practices of the large internet companies.

If applied to platforms with significant market power with respect to data sharing and data use, this principle would require the platforms to prove that they do not use the data collected on their business users to gain an unfair advantage vis-à-vis these business users. Also, according to some experts, rather than relying on antitrust law alone, a dedicated regulator for the tech industry should be created, to match those covering the banking, health and transportation sectors of the economy. A watchdog with expertise in the tech field could better review platform behavior and use of data on a case-by-case basis.

8.3. Overcoming technical barriers: data interoperability and portability

A right of data portability has been introduced with respect to personal data (Article 20 of the GDPR), and the platforms are currently working on technical interoperability solutions to implement it\textsuperscript{145}. The principle of data portability based on the GDPR is consumer-centric, allowing consumers to switch from one service provider to another, and move their data with them. However, the scope of this regulation is limited as it only applies to data that has been “provided” by the data subject to the data controller (i.e., it does not cover data uploaded by third parties (e.g., reviews, ratings) and inferred data. Also, the GDPR does not give a right to continuous access to data which would be necessary for synchronisation of accounts on different platforms or for providing data access to third-party service providers. Finally, the GDPR only applies to data relating to natural persons. Thus, data (including reputational data) relating to a legal person (in this case - platform business users) is not covered.

Various researchers and practitioners argue that similar business user-centric rights to the portability of non-personal data are necessary as well, to empower platform business users to ultimately decide whether their data is to be shared.

According to theoretical considerations, increased non-personal business data portability across platforms could enable new business models and facilitate market entry of new platform players. Data portability would also likely mitigate the lock-in and network effects of the dominant players by reducing user switching costs and allowing new market entrants to easier obtain critical mass of information. For example, a dealer of sneakers using Amazon to sell his products could decide to leave to the new marketplace “X-market”, and move the description of the products and comments of the buyers form the Amazon’s reputational feedback system.

However, as explained in Section 7.3., platforms have no incentives to develop costly interoperability and portability solutions. Given the expected positive effects of practical data portability for innovation and competition, intervention with a biding regulation could be considered. The proponents of this view argue that regulation to promote “true interoperability is vital as the market alone will always reject it as a threat to proprietary advantage”\textsuperscript{146}. They often use the examples of mandated portability in the banking and telecom\textsuperscript{147} sectors as a proof of how competition can be stimulated by regulation, despite a strong initial opposition by some companies in these sectors.

According to some authors,\textsuperscript{148} besides building on the Regulation on a framework for the free flow of non-personal data in the European Union\textsuperscript{149}, Article 9 of the P2B Regulation could be extended to include the non-personal data portability rights. However, to proceed further with mandating portability of non-personal (i.e. business user rather than consumer) data, solutions to several issues should be considered:

— The right of privacy of the consumers (of the business users services) may be compromised, as they will not know what the new data controller will do with their data after it is transferred. One solution could be a consent and opt-out mechanism (e.g., a consumer is asked for consent when leaving a review), provided by platform, for cases when a business user wishes to port its

\textsuperscript{145} E.g, the Data Transfer Project.
\textsuperscript{146} https://publications.parliament.uk/pa/ld201719/ldselect/ldcomuni/299/299.pdf
\textsuperscript{147} https://themarket.ch/interview/the-concentration-of-power-at-large-tech-companies-is-scary-ld.1290
\textsuperscript{149} https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018R1807
data to other platforms (e.g. in case a business user wants to export all data about his reputation). A data protection authority should oversee these processes and protect the consumer rights.

— To implement data portability in practice, industry players must work together to develop the necessary design, interoperability solutions and portability protocols. If the data portability standards are set by a small number of dominant players (as illustrated by the Data Transfer Project), smaller ones will be forced to follow this standard, carrying the costs of technical implementation that may be especially large for them.

— It is also important to define what specific data is relevant for portability, as different platforms have very different functionalities, use specific types of data which are not necessarily used in/by other platforms.

— The distinction between provided, observed and inferred data is very relevant. Provided data does not present particular difficulties for portability. However, sharing of inferred data (although it is built on the provided and observed user data) is especially tricky, as with the access to it, one can understand how the algorithm creating these data works. The algorithms concern intellectual property rights and are legitimate business secrets of the platforms. The solution may require a new framework for the protection of intellectual property concerning algorithms and Big Data, which should be designed to allow exchange of data without compromising business secrets150.

8.4. Important considerations

Several aspects should be carefully considered when exploring possible policy solutions to the issues related to platform data sharing and data openness, also in the light of the current COVID-19 pandemic.

First, technology evolves fast. At the current stage, the practices of data sharing in platform ecosystems are also evolving. In fact, data sharing is becoming more and more common, which is also illustrated by the platform case studies. New data sharing markets are developing as well, along with new business models and uses of data. Therefore, any new regulation must consider the need for flexibility in order to make sure that it remains relevant in changing technological in market contexts.

Second, questions of privacy, competition and consumer interests are strongly linked when the data sharing between platforms and their business users. Although not covered in in paper, the consumer perspective is crucial to understand in designing policy options. Platform data collection and sharing activities have very direct and potentially very strong (negative) impacts on privacy and personal data protection. Any normative instrument should seek an optimal balance between these interests.

Third, there is a risk that regulation might reinforce the existing dominant market positions151. Data ownership rights is an example. As demonstrated in our study, data scraping has become a crucial resource for many companies and other actors that use platform data152. If such data becomes subject to protection under intellectual property rights, this poses a risk of undermining the business model of data brokers, which would further increase the market power of large platforms153.

Moreover, any regulatory solutions should take into account the interests of small businesses. For instance, some smaller platforms expressed concern over the influence of big players (e.g., Amazon), which have become so influential in defining policy solutions that “ultimately only they can comply with”. Standards for data interoperability are also likely to be created by the major platforms without necessarily considering the smaller players. Meanwhile, small tech companies worry that opening up more consumer data - including consumers of their and competing products - would create an advantage for larger companies who have the resources to capitalise on it.

Finally, the COVID-19 crisis has been a major disruption in the platform economy, which has already led to notable innovation, as well as changes in the online markets. This also has implications in terms of data collection and data access. On the one hand, the coronavirus pandemic has been detrimental to many platforms, especially in tourism sector. For instance, Airbnb which had been one of the fastest growing online booking platforms, lost billions of USD in company value, and according to some opinions it may never recover to the pre-lockdown situation. Tripadvisor even discontinued its public policy operations in Europe. The reduction in the scope of platform activities in this industry, naturally, results in a reduction of data volumes generated, processed and shared. On the other hand, some platforms have started to apply increasingly intrusive measures to collect personal data on individuals. For instance, a data sharing deal between Apple and Google, as described in Section 5.3., is evaluated by some as a potential risk for individual privacy, other aspects of social life, and even democracy. Although the official reason for the expanded personal data collection activities remains the COVID-19 prevention (i.e. through contract tracing), these practices may not go away easily once the crisis is over. The policy solutions - both “soft” and “hard” - should take these and other emerging issues related into consideration as the situation develops. The platform economy post-COVID-19 may be quite different from that existing before.

---


9. Conclusions

The analytical paper built on the earlier analysis of platform data access\textsuperscript{156}, interviews, literature review and desk research to review the ecosystems of platform data flows in three sectors heavily intermediated two-sided platforms: e-commerce, mobile applications and tourism. First, compared to the earlier paper, the scope of analysis was extended. The comprehensive sector overviews were built on the evidence from 15 platform case studies in total, focusing on platform data access not only for their business users, but also for other third parties. Second, we also provide a more in-depth discussion on the main factors that have influenced the existing platform data sharing approaches, as well as a consideration of policy options to overcome the barriers for a wider platform data sharing. Third, another important achievement of this paper is the further development of the conceptualization and classification of data types relevant in platform data discussions. The new dimensions of this conceptualization involve the data subject (company’s own data (business or consumer) or competitor data), intended uses (re-use or provide a service to the platform) and purposes (upstream or downstream processes).

The paper looked specifically into the platform data generated through intermediating relationships and transactions between at least two sides of users: businesses and consumers. Moreover, we looked only into the data shared for the purposes of re-use by business users and other third parties - either for business decision making and improvement (downstream processes), or for developing new services and products with this data as an integral part (upstream processes). This left outside the scope of this specific paper the variety of intricate data sharing relationships between platforms and their business partners or service providers (e.g., technology solutions, payments, research, advertising) - at least to the extent that they are not allowed to re-use platform data for their own business purposes based on their bilateral agreements.

**BUSINESS USER ACCESS TO PLATFORM DATA**

In the e-commerce sector, marketplaces increasingly invest in the methods to provide useful data and analytics to their business users. Most of them, besides the data necessary to implement the transactions, offer their business users data on a number of metrics, covering their business results, as well as some information on their competitors. The analysis based on accessible evidence did not identify significant differences in the levels of business user data access provided by vertically integrated and non-integrated platforms. However, several merchants indicated that the data practices of the vertically-integrated marketplaces are concerning, as they have all the competitor data, while the sellers get almost none.

App stores are designed with the intention to share the data they generate as two-sided platforms with app developers - their business users - only. App developers, differently from business users of platforms in other sectors, also have significantly higher capacities to collect and analyse data themselves. Integral parts of the broader and very different ecosystems of Apple and Google, the app stores are both the gatekeepers and enablers of the third-party developers to access the customers of Apple’s and Google’s products. Their role in ensuring consumer data privacy is therefore very important both for the app store reputation and from the societal perspective, and acts as a limitation for giving a broader access to data for the developers. The data and analytics related content and functionalities provided by the two largest app stores to app developers differ rather marginally. Both App Store and Google Play offer information

\textsuperscript{156} Analytical paper #1: Platform data access and secondary data sources.
on the performance of the apps on their systems, while it is up to the developers to collect consumer permissions (within certain limits) and gather more specific app usage data.

The platforms in the **online tourism services sector** also maintain similar approaches to what data is and is not provided to properties using them. All the OTAs argue that the level of data they provide to their business users is sufficient or even superfluous to many groups of users. Given that the vast majority of properties on their platforms are small businesses, many of them do not fully exploit the possibilities provided by OTA analytics. Meanwhile, for the large hotel chains the analytics are not sufficient enough to be actionable, and they use own capacities to develop tailored insights.

Moreover, both small and large business users expressed that the key and most problematic access to data held by platforms concerns nominal information about their guests and their contact details. The OTAs holding the largest shares of the OTA market refuse to provide this information, causing difficulties for hotels to provide personalized services, identify undesirable reservations and maintain a customer relationship. Platforms, meanwhile, present a number of reasons why this data non-disclosure is justifiable, including protection of consumer privacy and protection of their business model - which brings a lot of value to hotels themselves.

In general, the amounts and types of data shared and not shared by platforms with their **business users** are quite similar across the three overviewed sectors. Most of the information they provide concern the performance of a specific business user on the platform. Data on competition or broader market overviews are provided in an aggregated form, with low granularity, and, as the business users claim, these are sometimes hardly actionable. Consumer personal data, including names and contact details is off limits for their business users in most cases, although this is the most desired type of data for advertising, direct consumer relationship and other purposes. The similarities of data provision by the platforms could be explained by another finding that the incentives and barriers that they face regarding the sharing of the data are similar as well.

Some of the data that platforms do not share or alternative analytics, e.g. related to the prices or competition in the market, can be collected from other sources or inferred. A notable segment of companies providing these services exists (described below in more detail). Other types of data, such as details of specific customers, are held only by the platforms and there are no substitute ways to acquire them for the business users.

It is also important to note that the types and amounts of data that platforms provide to their business users change rapidly. New indicators and metrics are introduced by platforms to their business users every year or more often. These changes, at least in the past several years, show a trend of increasing data access for business users. However, the amounts of data shared remain miniscule given the volumes of data that platforms collect.

**WIDER PLATFORM DATA ACCESS**

Although the business users are the primary intended recipients of the platform data, the ecosystems of platform data flows are not limited to business users as the sole type of businesses interested in them. This is despite that fact that platforms claim that they do not actively share their data with any other businesses for the purposes of re-use. It seems that ultimately the distinction with whom platform data is shared (e.g., business users, other businesses, other platforms) is less important. If a platform opens a data flow primarily intended to their business users (e.g., through their accounts) or customers (through the websites), it cannot ensure that other actors do not get access to this data as well.
To begin with, the paper has identified a large number of ancillary service providers for business users, specialized in marketplace optimization, insights and analytics, and offering data-as-a-service for the business users of overviewed platforms. Their activities require high volumes of platform-specific data. Although the platforms deny having direct data sharing relationships with them, they manage to gather the data mostly through web scraping (as a lot of relevant information is displayed publicly in on the platform websites) and getting it directly from the registered platform business users.

Further, the data generated by the two-sided platforms often have not only business but also societal and policy relevance. The best example is the short-term rental market, emergence of which was enabled by the rising platform economy. It has led to broad consequences in the real estate markets and in life in European cities more generally. Because of this, data held by accommodation platforms have become of interest for researchers, NGOs, public authorities and policy makers. The primary method to acquire it remains web-scraping. This is enabled by the fact that platforms must display a lot of information about their listings on their websites for consumers.

On another note, all platforms use services of various third parties, including, for example, research, consultancy, marketing and advertising, payments and so on. To provide the needed services for the platforms, the third parties must receive certain types and amounts of data held by the platforms. The contracts between them usually are very specific about the purposes that this data can and cannot be used, and any kind of data re-use for other purposes than providing a specific service for the platform are prohibited. However, the position of Google and Facebook in the advertising market enables them to enforce their standard terms and conditions, which do not prohibit them from using data of other platforms for whatever purposes they wish to. This poses the risk for the platforms, given that both Google and Facebook are developing their own vertical businesses in travel, commerce, and other sectors.

INCENTIVES AND CONSTRAINS FOR PLATFORM DATA SHARING

The conclusion on the level of platform data access and is sufficiency is difficult to make. On the one hand, given the imperative to protect consumer privacy and own business interests, the current level of platform data access to their business users and other economic actors is optimal. Platforms do provide the minimum level of data needed by business users to process the transactions, manage business process, as well as to make more insightful business decisions, while consumer privacy is protected Small companies, especially in the tech sector, also appreciate that platforms protect the data (i.e. data that they contribute to by using platform services) from larger competitors. However, other businesses, especially large players in their respective markets, often complain about the lack of platform data access. Platforms keep themselves enough exclusive data, which enables them to concentrate market power if used in certain ways.

The current level of data access provided for platform business users can be viewed as shaped by reconciliation of several tensions that these platforms face, including economic, legal and technical factors:

- In providing consumer and behavioural data to developers, app stores seem to be balancing data protection regulation and their general aim that their developers succeed in innovating and putting high-quality apps on their marketplaces, therefore generating more transactions.
- OTAs face similar trade-offs what concerns consumer behaviour data and analytics. However, given that one of the platform goals is to maximise the number of transactions conducted using them, some business users suspect that some of the analytics they provide may be biased. When it comes to nominal and contact data of consumers, OTAs also balance between the need to ensure successful stays at the hotels and prevent hotels from bypassing their platforms to arrange bookings.
In e-commerce, the same consumer privacy versus success of the business users/ volume of transactions dilemma is common for most marketplaces. In addition to this, even though the evidence from case studies is not sufficient to claim that vertically-integrated platforms systematically tend to provide less data to its business users, they also may have incentives to take the success of their verticals into the equation.

In addition to these, issues of technical implementation, investment in data processing and others concern most platforms. They tend to invest into these only as much as this can pay-off relatively quickly. Because of this, technical solutions from interoperability and portability are not likely to be developed as a result or market forces only. Meanwhile, the core role of data in platform business models and competition becomes important regarding the incentives and constraints for general openness of platform data.

POSSIBLE SOLUTIONS

The analysis identified several more specific issues that policy solutions could address to enable greater openness of platform data and to reduce the possibilities of its anti-competitive use.

First, companies in the tech sector - including platforms - have built their businesses around data. Competition between the main platforms is fierce, and given that data is at the core of their business models and success recipes, they aim to keep most of it under locks. Platforms also cannot trust that opening more data would not turn against them. For example, collected evidence shows that at least some American and European platforms feel being increasingly threatened by Chinese competitors, leading to even more cautiousness in opening more data. Any binding regulation mandating data openness should consider in detail what data should be shared under what circumstances, by whom and to whom, in order to protect the interests and innovation potential of European companies. A lot of additional in-depth research is needed to develop specific legislative proposals.

Second, the interview programme and analysis of other evidence showed that the key competition-related stemming from data advantage concerns vertically integrated platforms. Business users in e-commerce, as well as OTAs (as business users of Google) expressed concerns that vertically integrated platforms combine data collected on them, their consumers and business users to develop competing products. One of the possible approaches to overcome this is limiting the possibilities of these gatekeeper platforms to use the data that other businesses submit them for purposes other than providing services for these businesses (e.g. for developing their own business verticals, re-personalising consumer data, etc.). Given that practically this would be very difficult to detect, the principle of reversing the burden of proof might be considered, along with an establishment of a watchdog agency overseeing digital platform activities.

Finally, in the eyes of some business users, doing business with the major platforms dominating specific markets is inevitable. This is related to the strong network effects that these platforms demonstrate and the lack of alternatives. According to some experts, a possibility of business users to port their non-personal data between different platforms could encourage innovation and reduce the entry costs for new platforms, then leading to an increase of market players and decrease of market concentration (by reducing business user switching costs and therefore platform network effects). However, platforms have no incentives to develop interoperability solutions for this to become a reality. On the contrary platforms are likely to strongly oppose such proposals for developing data portability functionalities, which are not only costly, but also potentially threatening to their market positions based on the control of exclusive data. Some researchers argue that mandating data portability for business user data could be a solution for this, leading to fairer competition in the market - as showed by the telecom and banking sectors in which such regulations were introduced.
Nevertheless, several aspects should be carefully considered in addressing platform data access with policy measures. First, technology evolves fast. At the current stage, the practices of data sharing in platform ecosystems are also evolving. Therefore, it seems sensible not to intervene with binding regulation for data openness and the need of flexibility should be kept in mind, to allow the legislation to remain relevant in new technological in market contexts. Second, questions of privacy, competition and consumer interests are strongly linked when the data sharing between platforms and their business users is in question. Any normative instrument should seek an optimal balance between these. Although not covered in in paper, the consumer perspective is crucial to understand in designing policy options. Third, excessive or not sufficiently tailored regulation might reinforce and confirm existing dominant positions\(^\text{157}\), resulting in even more market concentration. Fourth, any intervention should consider the effects on both larger and smaller industry players, and ensure that policies do not leave the latter in a disadvantaged position. Finally, the COVID-19 crisis presented an unprecedented shock to many sectors of the platform economy, extending to data collection and sharing practices as well. While at the time of writing of this report the situation is still developing, the policy discussions should take the emerging issues into account.

Annex 1. The case studies

Amazon

Aside from being the largest online retailer in the U.S., Amazon is one of the biggest e-commerce companies in Europe. Among a broad range of other services, it also operates one of the world’s top marketplaces with over 2 billion visits to Amazon.com alone every month and $160 billion worth of global third-party sales in 2018 (compared to $95 billion for eBay).

Amazon is reported in the media as the largest retail data repository in the world. It has a lot of information on the consumers and business users active on its marketplace (e.g. browsing, purchase, selling activities; how, where and when Amazon services are consumed). The company claims to collect personal information through three main channels: data provided by users, data collected automatically and data received from third-party sources. A summary of Amazon’s data collection is provided in the box below.

BOX 5. DATA COLLECTED BY AMAZON

- Information provided by buyers and sellers:
  - By providing or configuring settings, account or profile information, customers reveal personally identifiable information such as name, address, phone number, email address, age, personal description and photograph, signature, payment information, privacy preferences, IP address. Sellers reveal their product information, as well as business demographics (sector, business type, registration address, year of business and number of employees).
  - Corporate or financial information
  - Credit history information
  - Device information
  - Voice recordings (when speaking to Amazon’s Alexa virtual assistant)
  - Images and videos
  - Information and documents regarding identity, such as Social Security or driver’s license numbers
  - Content of reviews, calls and e-mails sent to Amazon (when communicating with the platform through these channels)
  - E-mail addresses of friends and other people
  - Location data and information on the other apps and services used by the customer (when consumers are downloading, streaming, viewing, or using digital content on a device, or through an app).
- Information collected automatically:

---

160 Amazon. (2018). Annual report. Available at https://ir.aboutamazon.com/static-files/0f9e36b1-7e1e-4b52-be17-145dc0d8b5ec
162 Amazon. (n.d.). What data does Amazon collect and use? Available at https://www.amazon.co.uk/gp/help/customer/display.html?nodeId=G6RZ4RMNMLUQ8L2Y2
<table>
<thead>
<tr>
<th>User interaction with content and services available through Amazon Services, e.g. content downloads, streams, and playback details, purchase and content use history (which is sometimes aggregated with similar information from other customers to create features such as Best Sellers), click-stream data (to, through and from Amazon websites, including data and time), product and content views and searches, length of visits on pages, interaction information (e.g. scrolling, clicks and mouse-overs), other browsing and usage data.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

- **Information received from third parties:**
  - Updated delivery and address information from carriers and other third parties
  - Account information, purchase and redemption information, and page-view information from some merchants with which the company operates co-branded business or for which technical, fulfilment, advertising or other services are provided by Amazon.
  - Search results and links
  - Credit history information from credit agencies
  - Information about internet-connected devices and services linked to Alexa

---

In addition to the extent of data collection described officially in Amazon Privacy Notice, interviews with Amazon business users and desk research revealed that the company has a very fine-grained information on third-party seller activities, such as click-through rates of each listing, paths through which customers end-up viewing specific products, complementarity of products and price level adjustments, and very detailed information on individual customers and their shopping habits. To better understand customers, the platform also uses external datasets, such as census data for gathering demographic details. This data can also be used to develop very detailed profiling of each customer.

There is also evidence on the platform’s investments in expensive experiments for data collection. According to some reports, Amazon runs experiments that involve funding discounts on third-party products in order to gather information on consumer price sensitivity. Despite the anecdotal evidence coming from whistle-blowers, such as the company’s ex-employees, the full extent of the customer data collection by Amazon and methodologies applied is not publicly disclosed.

Nonetheless, linking data from these various streams, primarily business users’ data with customer data, results in a wealth of information on how the market functions, and presents the company with basically unlimited possibilities to analyse this data to draw commercially-viable insights. Through advanced analytics, Amazon has a clear understanding of what people actually buy, search for and do not buy, and what they may buy next - this gives it an extreme competitive advantage.

---


166 Financial times. (n.d.). Amazon’s ever-increasing power unnerves vendors. Available at https://www.ft.com/content/c82ce968-bc8a-11e8-94b2-17176fb93f5


168 Financial Times. (n.d.). “Amazon’s ever-increasing power unnerves vendors”, Financial Times. Available at https://www.ft.com/content/c82ce968-bc8a-11e8-94b2-17176fb93f5
Speaking about Amazon Marketplace data access to third parties, business users (sellers) and vendors are the primary intended recipients. We cover the company’s data provision to its third-party sellers in the following section.

Aside from business users, Amazon also shares consumer’s personal information with Amazon.com, Inc. subsidiaries and other third-parties, such as third-party businesses that are involved in customer transactions, and third-party service providers to Amazon (e.g. delivery companies, payment processing providers, credit scoring agencies, analytics and marketing providers). The platform also provides ad companies with non-directly identifiable information of its consumers for interest-based advertising purposes, and some third-party advertising partners may also collect information about Amazon Services users when they interact with their content, advertising, and services\footnote{Amazon (2020). Amazon Privacy Notice. Available at https://www.amazon.com/gp/help/customer/display.html?nodeId=2019090104GUID-1B2BDAD4-7ACF-4D7A-8608-CBA6EA897FD3\_SECTION\_87C837F9CCD84769B4AE2EBB14AF4F01}. However, this data is not shared for reuse purposes.

**Data provided to Amazon business users**

Only a small fraction of data collected by Amazon is shared with its third-party sellers, especially those who do not use Amazon’s additional services, such as fulfilment and advertising. According to the interviewed Amazon sellers, this becomes even more of an issue because of the platform’s vertical integration: Amazon Retail – often their direct competitors on Amazon Marketplace – knows everything about its competition and its competitors.

For third-party sellers who own a brand and are enrolled in Amazon Brand Registry (200,000 worldwide as of 2019)\footnote{Amazon Progress Report. Available at https://brandservices.amazon.com/progressreport}, Amazon offers several types of data and analytics. When a consumer purchases goods online through Amazon Marketplace, not only Amazon but also the third-party seller can acquire some consumer data (e.g. address or phone number to fulfil orders). However, there is a difference in terms of the degree and detail of the data. A seller can get the data displayed in the transaction, while Amazon obtains consumer data beyond the transaction data, including browsing history and clickstreams\footnote{Li, W. C., Nirei, M., & Yamana, K. (2019). Value of data: There’s no such thing as a free lunch in the digital economy. Discussion papers 19022, Research Institute of Economy, Trade and Industry (RIETI). Available at https://www.bea.gov/system/files/papers/20190220ValueofDataLiNireiYamanaforBEAworkingpaper.pdf}.

In addition, a seller on Amazon get access to some conversion metrics – essential for improving content on their product pages. Sellers can access information on their business performance (through Seller Central or an API\footnote{Amazon MWS Sellers API section of the Amazon Marketplace Web Service (Amazon MWS).}), such as:

- Various reports that help sellers manage their “Sell on Amazon” business, including Brand Dashboard, which provides suggestions what can be optimised, customer reviews (”Voice of a customer”);
- Business reports, including listing traffic, conversion rates, information on the seller’s accounts, shipping, fulfilment, finances, orders, invoicing and products;
- Amazon Selling Coach recommendations (representing the “insights” information category) for the following categories:
  - Advertising: Recommendations for advertising your products and for using the sponsored products program.
Fulfilment: Recommendations for the top products to fulfill through Fulfilment by Amazon (FBA).

Global Selling: Recommendations for expanding your products to more regions and marketplaces.

Inventory: Recommendations for restocking low or out-of-stock items in your inventory.

Pricing: Recommendations to review pricing on items in your inventory where your offer is not the lowest price.\(^{173}\)

Furthermore, since 2019, Brand Analytics data is provided to brand-registered sellers at an aggregated level (see table below). This data provides a valuable but rather broad overview of the relevant market segments and competition on the marketplace.

### BOX 6. AMAZON BRAND ANALYTICS, AS ACCESSED BY A SELLER

#### Amazon Search Terms
- For each search term:
  - department (.de, .co.uk, .com, etc);
  - search frequency rank;
  - #1, #2 and #3 most clicked product (by Amazon Standard Identification Number - ASIN) for that search term, with its click share (%) and conversion rate (%).
  - No actual search volume or volume of purchases in absolute terms.

- They can be filtered by reporting range (e.g., weekly, specific week).
- Could be downloaded in excel or CSV or explored in a web browser.

#### Market Basket Analytics
- For each of the seller’s products:
  - #1, #2 and #3 most purchased products on Amazon, purchased together with that specific product of the seller, with its combination percentage (i.e., what percentage of customers, who bought a seller’s product, also bought other top 3 products).
  - Can be filtered by category, subcategory, brand (among the seller’s brands), reporting range.
  - Could be downloaded in excel or CSV or explored in a web browser.
  - No actual volume of purchases.

#### Item Comparison and Alternative Purchase Behaviour
- For each product most frequently viewed by customers together with (i.e., the same day as) the seller’s products, in addition to customer’s final choice after viewing seller’s product:
  - #1, #2, #3, #4 and #5 compared products titles and ASIN (sold by competing sellers), with compared percentage (% of times this alternative product was viewed by customers who also viewed the seller’s product within the same day).

---

\(^{173}\) Amazon Services Europe. What you should know about the Amazon MWS Reports API section. Available at http://docs.developer.amazonservices.com/en_UK/reports/Reports_Overview.html
Businesses using Amazon’s paid advertising tool can get additional data and analytics, such as ACoS (advertising cost of sales), showing the ration between advertising costs and sales revenues. In autumn 2019, the platform also launched Amazon Attribution (beta), which is available to sellers enrolled in Amazon Brand Registry, vendors, and agencies that advertise to sell products on Amazon. It is a measurement tool that tracks the performance of non-Amazon media (such as ads, social ads, and display ads) and how it helps to drive sales to Amazon. The attribution measures traffic (impressions and clicks) conversion, including sales driven by external campaigns. Moreover, sellers can also advertise on Amazon through Amazon Advertising, which gives them access to data on their ads performance.

According to the majority of interviewed Amazon sellers and desk research, information provided by the platform on an individual seller’s operation and performance is useful and sufficient for simple insights. Brand Analytics is also appreciated by the seller community as it is free, allows to “keep an eye on the competitors” (at least to some extent), helps to apply potential keywords, and provides better knowledge of the target audience and products. Nonetheless, some interviewed sellers argued that Brand Analytics only gives a minuscule part of the full picture: the extent, granularity and comprehensiveness of data shared with sellers is still very limited. Some interviewees argued that brand Analytics is simply not enough considering what access to data Amazon Retail has.

To compensate for the lack of information on the marketplace trends and activities, most of the interviewed merchants widely use services by online marketplace optimisation tool providers, which we cover in the following section, or scrape Amazon data available publicly. This is not surprising given that 1) Amazon does not allow to integrate tools such as Google Analytics with Amazon seller accounts, which would allow to gather additional insights, and 2) Amazon sellers are very eager to expand their business (97% of sellers plan to do so) and optimise current product listings (46%), in which data plays a key role.

In this context, Amazon third-party sellers are not on the level playing field with Amazon private label brands: the latter have “all the data in the world to know what products to create and exactly what keywords to target”. Some sellers noted the platform has all the needed means and information (from the manufacturer, logistics, to customer data) to remove and replace any of its successful sellers with own private label products (though this does not mean that it would have a reason to do this). A number of

---

https://advertising.amazon.com/amazon-attribution


https://www.practicalecommerce.com/using-amazon-ad-attribution-to-track-offsite-campaigns

https://www.junglescout.com/amazon-ad-attribution/

https://www.sellerapp.com/blog/amazon-brand-analytics/

https://www.sellersessions.com/amazon-seller-data/
stories on the media\textsuperscript{181} and testimonies by Amazon’s former employees\textsuperscript{182} shed the light on Amazon’s use of the aggregated seller data to boost Amazon retail sales and the resulting unfair competition between Amazon retail and third-party marketplace sellers. In November 2019, Amazon itself confirmed that it uses “aggregated data” from sellers in its third-party marketplace to improve its overall business\textsuperscript{183}.

In fact, Amazon’s dual role as a retailer and a marketplace, and whether it is using third-party seller data to enhance its retail offering is under an anti-trust investigation by the European Commission. In parallel, the company is also subject to ongoing or closed probes in a number of EU Member States (Germany, Austria, Luxembourg, and Italy) and is under similar anti-competition investigations in the US. The US investigations follow the Wall Street Journal reports on Amazon employees using data about independent sellers to develop competing products as part of its retail\textsuperscript{184}. As a platform in a dominant position (in terms of its reach; around 200 million customers each month\textsuperscript{185}), Amazon can use the data to identify successful products with a corresponding price, then to offer them itself, at a better price or delivery conditions.

The platform itself claims that it has no incentive to harm its third-party sellers, arguing that they generate an important and increasing share of its revenues (third-party sales amounted $160 billion globally in 2018 and accounted for 58\% of physical gross merchandise sales that year\textsuperscript{186}). On the contrary, Amazon claims to be investing in tools to help sellers scale their business.

According to Amazon representative, Amazon’s private label products generate approximately 1\% of Amazon’s total sales (this information is difficult to verify, as it is not provided in the company’s Annual Report; some reports provide the same figures as Amazon\textsuperscript{187}, others estimate that the private label business totalled in $7.5 billion in 2019\textsuperscript{188} or around 5\% as compared to the total turnover), and this is far less than other online retailers, many of whom have private label products that represent over a quarter of their sales. Also, not all private brands are successful\textsuperscript{189}. However, some sources estimate that Amazon’s overall private brands sales grew over 2\% from 2017 to 2018, and some segments grew significantly more (e.g., 81\% growth of private label consumer packaged goods)\textsuperscript{190}. Meanwhile, an American investment


\textsuperscript{185} Dayton, E. (2020), „Amazon statistics you should know: opportunities to make the most of America’s top online marketplace“ \textit{Big Commerce}. Available at https://www.bigcommerce.com/blog/amazon-statistics/a-shopping-experience-beyond-compare

\textsuperscript{186} Amazon. (2018). Annual report. Available at https://ir.aboutamazon.com/static-files/09be58b1-7e1e-4b52-be17-145d9d9b5ec

\textsuperscript{187} Green, D. (2019), “Amazon says its private labels are only 1\% of its business, but new data shows some are seeing huge growth”, \textit{Business Insider}. Available at https://www.businessinsider.com/amazon-private-labels-some-grow-quickly-data-shows-2019-4


\textsuperscript{190} Green, D. (2019), “Amazon says its private labels are only 1\% of its business, but new data shows some are seeing huge growth”, \textit{Business Insider}. Available at https://www.businessinsider.com/amazon-private-labels-some-grow-quickly-data-shows-2019-4
bank projects that Amazon’s private label business could see revenues of $25 billion by 2022. The data that Amazon collects on all the transaction on the Marketplace contributes significantly to this.

We broadly summarise sellers’ access to data and analytics provided by Amazon in Table 3 table below.

### Table 3. Types of Data and Analytics Accessible by Amazon Marketplace Sellers

<table>
<thead>
<tr>
<th>Amazon</th>
<th>User Identification Details</th>
<th>Data on Transactions Between Businesses and Customers</th>
<th>Business Performance</th>
<th>User Behaviour</th>
<th>Analyses of Market Trends/Developments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Own</td>
<td>Own/Competitor</td>
<td>Own/Competitor</td>
<td>Own/Competitor</td>
<td>Own/Competitor</td>
</tr>
<tr>
<td>Raw data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analytics</td>
<td>NA</td>
<td>Some</td>
<td>Yes</td>
<td>No</td>
<td>No/No/NA</td>
</tr>
<tr>
<td>Insights/Actionable guidelines</td>
<td>NA</td>
<td>Some</td>
<td>Yes</td>
<td>Some</td>
<td>Some/Some/Some</td>
</tr>
</tbody>
</table>

Data provided to Amazon Vendors

Vendors are the businesses that sell wholesale to Amazon Retail which operates on the Amazon Marketplace. Vendor relationship with Amazon is somewhat similar to that between a supermarket and its suppliers. Until recently, Amazon vendors had two options: accessing only elementary data (such as sales and inventory) through Amazon Retail Analytics Basic (ARA) package, or paying reported $30,000 a year or more for access to Amazon Retail Analytics Premium (ARA Premium), which offered more in-depth data. Access to ARA Premium allowed vendors to get more, but not all, of the data that Amazon has compiled. Amazon’s position is that they, not the vendor, ship the product and own the customer relationship, so they do not give anyone customer profiles, email addresses or allow contact. There have been reports that some larger vendors may be provided with more information on a case-by-case basis and that access to various ARA reports depends on permission and the contract terms of the vendor agreement. One interviewee owning several large businesses active on Amazon added that vendors would even register as third-party sellers on Amazon to get access to more data.

However, since April 2020, Amazon has officially rolled out its new reporting tool for Vendor Central, Brand Analytics, which mimics the roll-out of Brand Analytics to third-party sellers earlier in 2019.

---

Brand Analytics replaces ARA Premium, and ARA Basic for those who are Brand Registered, through automatic upgrades. Reportedly, all the same data is available under Brand Analytics, as was under ARA Premium (see FIGURE 6 figure below). Vendors who are not brand owners, or who are brand owners but are not enrolled in Amazon Brand Registry, will retain access to ARA Basic. This Amazon’s move is said to be consistent with Amazon’s recent efforts to make selling more favourable to brand owners (arguably, in order to address the counterfeit problems), and gives brands the long wished-for insights into customer behaviour. It certainly makes a major shift in the importance of being brand-registered with Amazon.

**FIGURE 6. ARA PREMIUM VS BRAND ANALYTICS**

![ARA Premium vs Brand Analytics](image)


**Data provided to other entities/ open for collection**

As the largest global e-commerce marketplace, Amazon is targeted by the largest number of data insights providers and OMOTs. Amazon’s representative estimated that around 1,700 data and software companies serve the sellers of Amazon alone, offering various optimisation products based on listing, business user, keyword and sales analyses. The platform claims to have no direct relationship or data sharing arrangements with these players. However, a lot of data is publicly available. For example, offer prices, listing descriptions, ratings, reviews, and relative popularity of an item are visible on the platform’s websites. Amazon also publishes lists based on unit sales data, like the Best Sellers, New Releases, Movers

---

& Shakers, Most Wished For (items that consumers add to their wish lists), Gift Ideas\textsuperscript{199}, as well as ASIN pages, that show sales rank of a product in a specific product category. While these lists are primarily aimed at customers, sellers and other third parties also make significant use of this data, according to a company representative.

The primary way to get this data into useful datasets is web scraping. In fact, Amazon shared that an estimated 35\% of their traffic comes from automated data collection tools. Examples of third-parties using scraped data include AMZ tracker (tracks competitors sales and rankings, and provides reports) CamelCamelCamel (a price comparison tool) and Keepa (another price comparison tool). According to a company representative, they all use automated web data gathering tools. Despite scraping (for commercial purposes) being against Amazon Terms and Conditions, the marketplace has certain tolerance since these tools are beneficial for consumers and as long as they do not interfere with Amazon’s business. However, this tolerance has its limits; for example, Amazon has banned several entities that were acquiring the user data from the platform in violation of its policies concerning consumer privacy\textsuperscript{200}.

Moreover, entities in advertising relationships with Amazon may get some additional data. For example, Camelcamelcamel (offering a price tracker service that provides price drop alerts and price history charts for products sold Amazon\textsuperscript{201}, meaning they have historical pricing and sales rank data\textsuperscript{202}) is on Amazon Associates Programme (Amazon Services LLC Associates Program). It is an affiliate advertising programme enabling sites to earn from affiliate marketing linking to Amazon.com. Affiliates can setup access to Amazon’s Product Advertising API\textsuperscript{203} through which they ‘must be able to fetch out product content such as prices’\textsuperscript{204}. Amazon is reported to funnel sales data directly to Camel\textsuperscript{205} although on their website, the Camel do not provide information on how they get data\textsuperscript{206}. This cooperation is claimed to be Amazon’s attempt in satisfying the most persistent deal-hunting consumers while keeping the differentiated/dynamic pricing strategy\textsuperscript{207}, which has resulted in a public outcry in 2000. In an interview the Camel founders expressed their fears of being cut-off from data supply by Amazon despite them being cooperative\textsuperscript{208}. Reportedly, similar Amazon price tracker sites such as Keepa.com, Tretracktor.com, mypricetrack.com and Unimerc.com work on the same principle – being fed price information directly from Amazon as affiliates\textsuperscript{209}.

Overall, this whole boosting ecosystem of businesses thriving on access to Amazon’s public data is heavily reliant on it. For instance, a free tool Yasiv used by Amazon sellers, that allowed to check the popularity of and connections between the products on Amazon based on consumer purchases, has recently stopped working due to Amazon’s updates on its API, which no longer allows to fetch related product data (see Figure 7 below).

\textsuperscript{199} Amazon (n.d.). Amazon Best Sellers. Amazon. Available at https://www.amazon.com/gp/bestsellers


\textsuperscript{201} https://uk.camelcamelcamel.com/about

\textsuperscript{202} This is evident on their own website: https://uk.camelcamelcamel.com/about

\textsuperscript{203} https://docs.aws.amazon.com/AWSECommerceService/latest/DG/Welcome.html

\textsuperscript{204} https://www.quora.com/How-does-CamelCamelCamel-scrape-the-prices-of-Amazon-s-products

\textsuperscript{205} Hutton, R. (2013) “Amazon Discount Tracker Camelcamelcamel Tips Users to Deals”, Bloomberg. Available at: https://www.bloomberg.com/gp/best/vol-71/12-05/amazon-discount-tracker-camelcamelcamel-tips-users-to-deals-

\textsuperscript{206} https://camelcamelcamel.com/blog/how-our-price-checking-system-works?


\textsuperscript{208} Ibid.

Besides the data displayed on Amazon's websites, a lot of open data available on Amazon Web Services. For example, the Amazon Customer Reviews dataset (also known as Product Reviews) provides historic data since 1995\(^\text{210}\). The dataset is intended for academic research into the fields, such as Natural Language Processing (NLP), Information Retrieval (IR), and Machine Learning (ML), which can facilitate study into customer product experiences and their evaluations.

There are also some accounts on Amazon's involvement in secret data sharing deals with the other big tech companies, but the evidence is limited. For example, in 2018, The New York Times reported on the internal Facebook documents on special data (including personal data) sharing arrangements with over 150 companies\(^\text{211}\). Amazon was among the key partners and one of the companies that benefited from access to users' names and contact information through their friends, and which in turn, was reported to provide Facebook with contact lists. However, the validity of these claims cannot be checked.

### Allegro

Allegro is the largest third-party marketplace in Poland with 17 million active users visiting the site every month, over 100,000 sellers (mainly Polish SMEs) and 1.5 million transactions a day\(^\text{212}\). According to a company representative, over 99% of its revenues come from merchants (third-party sellers) and less than 1% from Allegro selling directly\(^\text{213}\). Allegro’s main revenue streams are transaction fees charged to merchants (there’s no entry or listing fees) and advertising\(^\text{214}\). In terms of the European scope, Allegro is the sixth largest EU’s online retailer\(^\text{215}\). The marketplace is similar to eBay as it is a pure intermediation service provider, not engaged in warehousing or logistics for merchants.

---


\(^{212}\) [https://raportcsr.allegro.pl/en/about/company](https://raportcsr.allegro.pl/en/about/company)

\(^{213}\) Although the Official Allegro Shop generates less than 1% of the total sales from the platform, Polish Competition Authority is investigating whether Allegro abused its dominant position by favouring its own sales over sales of other retailers on the marketplace: [https://www.lexology.com/library/detail.aspx?g=bdfac899-5538-4682-84fd-19b4feb8205](https://www.lexology.com/library/detail.aspx?g=bdfac899-5538-4682-84fd-19b4feb8205)


There are two account types on Allegro: standard and business. Standard account is for private sellers (individuals), while the Business Account is for companies/ professional sellers (Konto Firma). To help non-professional sellers, in 2019 Allegro launched Allegro Lokalnie (“Allegro Locally”), which is meant to eventually become a marketplace for private sellers (C2C platform) while Allegro.pl (main B2C platform) will remain for companies only\(^{218}\). Only private persons who have a standard activated account at Allegro.pl are encouraged to sell on Allegro Lokalnie, where they are expected to migrate to from the main platform by 30 June 2020\(^{217}\). At the same time, businesses running on a standard account on Allegro’s main platform are advised to convert to a business one\(^{218}\). In practice, despite the launch of Allegro Lokalnie, there is no 100% split between the B2C and C2C platforms as they both can be used by standard and professional account holders.

Data is an important asset for the company, and it is seen as a competitive advantage against global players. Therefore, Allegro is investing increasingly into the processing of data for the benefit of offering better services to its users. The company noted a huge potential in data and willingness to build and use it more (e.g., in enhancing machine learning capacities). Like the other marketplaces, Allegro collects both personally identifiable information (PII) and non-personal information through various channels, as summarised in the box below.

**BOX 7. DATA THAT ALLEGRO COLLECTS**

- Email address, passwords, date of birth, contact details, data about the company, which is provided directly by its users when setting up an account, carrying out activities or transactions on Allegro, contacting the user service on the platform.
- Automatically collected data from the user’s browser and device, such as cookies, IP address, software and hardware parameters, webpages viewed, mobile device identification number and other data regarding devices and the use of systems.
- Data on user activities (searched keywords, posted listings), sales, trends, number of transactions, advertising (what Allegro calls 'data efficiency campaign', e.g. advertising that merchants are pursuing).
- Data obtained through technologies such as cookies and during the use of Allegro mobile application (location of user device). Among the cookies used are the analyses and studies ones, which allow to track the website activity of the user and understand their preferences, improve and develop products. Typically, cookies do not constitute personal data unless combined with other information.

Source: compiled by PPMI, based on Allegro.pl Privacy Protection Policy\(^{219}\) and interviews

When it comes to data sharing with third parties, the company advocates for voluntary data sharing that ensure proportionality and safeguards, and allows the company to assess and decide whether to open their data on a case by case basis.

Since Allegro is a pure marketplace, it has partners that offer solutions such as software and other services: personalised email communication, cloud infrastructure, data storage, advertising and content solutions, financial, delivery and analytical services and other. Allegro shares data with these third-party entities, which enables the company to offer full customer. This includes personal data, but service providers are not authorised to use this data for own purposes (i.e. re-use it), only in order to carry out services for or on behalf of Allegro.

---


---

73
Data provided to Allegro business users

In terms of business user access to data, it differs by subscription type. For additional fixed fees\(^{220}\), sellers can subscribe to Allegro “Abonament” (Subscription) basic or upgraded versions, which give access to Trade Analytics, which is a sales monitoring tool for the Allegro marketplace, based on data directly from Allegro.pl (sourced via Allegro API)\(^{221}\). Sellers with basic subscription get access to the basic plan of Trade Analytics, which offers access to limited sales data, while sellers with professional and expert subscriptions can access extended sales statistics in the premium package of Trade Analytics.

Trade Analytics provides information on sales volumes of different products, advice/insights to sellers on where they are missing an opportunity, suggestions and recommendations on what the seller could do to improve their overall performance, which products should be changed (see the box below for more detailed information). Subscribed sellers can also use an API connection for easier integration of Allegro’s and sellers’ own software, which also allows to get more data. Through its API, Allegro provides all the data on transactions in real-time, including individual-level anonymised data on all the sellers.

**BOX 8. DATA PROVIDED BY ALLEGRO IN TRADE ANALYTICS TOOL**\(^{222}\)

- Main report: statistics on offers such as sales volumes, list of products, types of promotions, shipping options etc.
- Offer details: sales value, number of units sold, page views, duration, competitor name, number of transactions, costs (listings, commissions, promotion).
- Price monitor: price changes on specific offers, including changes in pricing of competitors offers.
- Competition report: value of sales, number of transactions, position in the seller ranking, market share of sales in given category.
- List of offers: historical offer data on issued, ending, renewed offers and those occurring before, after or during a specified time interval.
- In-depth category report: data on changes in buyer habits depending on month or season, trends surrounding key holidays and promotional events (which categories are most popular).
- Allegro hits: top 1000 items and sellers throughout Allegro regardless of category.
- Top 100 in category: top 100 products in any category, generating the most sales and most often chosen by consumers. Also information on average prices and most popular keywords used in item descriptions.

Data on competitors is provided in an aggregated form, with confidential and anonymised company names. In terms of customer data, Allegro only shares what is necessary for delivery purposes: the seller receives contact information of the buyer currently appearing in their account settings, i.e. name and surname, e-mail address (encrypted by Allegro), telephone number and address\(^{223}\). These details are shared in compliance with the GDPR, according to the platform. Generally, access to consumer data for sellers is restricted to secure Allegro’s business model in making sure transactions happen on Allegro.

---

\(^{220}\) Basic subscription is provided for a fixed fee charged every 30 days (around EUR 9); professional subscription – for around EUR 44 fee charged every 30 days; expert subscription – for EUR 660 fee per 30 days. See [https://allegro.pl/pomoc/dla-sprzedajacych/abc-sprzedazy/abonament-co-ofera-i-dlaczego-warto-z-niego-skorzystac-a1ydPX6BwIx](https://allegro.pl/pomoc/dla-sprzedajacych/abc-sprzedazy/abonament-co-ofera-i-dlaczego-warto-z-niego-skorzystac-a1ydPX6BwIx).

\(^{221}\) [https://tradeanalytics.pl/about](https://tradeanalytics.pl/about)

\(^{222}\) Based on information provided on Allegro’s official website ([https://allegro.pl/pomoc/dla-sprzedajacych/abc-sprzedazy/statystyki-sprzedazy-w-abonamentach-allegro-ID3Kk05Egsn](https://allegro.pl/pomoc/dla-sprzedajacych/abc-sprzedazy/statystyki-sprzedazy-w-abonamentach-allegro-ID3Kk05Egsn)), Trade Analytics tool website ([https://tradeanalytics.pl/about](https://tradeanalytics.pl/about)) and collected during interviews

\(^{223}\) Article 7.4 of the Allegro’s Terms and Conditions. Available at [https://allegro.pl/regulamin/en#article-7-contract-concluded-as-part-of-the-listing](https://allegro.pl/regulamin/en#article-7-contract-concluded-as-part-of-the-listing)
According to platform representatives, Allegro’s position is that sharing unstructured data would not bring value to sellers as most of them focus on everyday business rather than on in-depth analysis on market changes. A platform is better placed to pursue such analysis and share meaningful data and insights with the sellers. As observed by a platform representative, seller may not have the capacity to interpret raw data even if the platform shared it and could ignore some of the data if everything was shared. Furthermore, sharing all the raw data with sellers could raise intellectual property rights (IPR) and security concerns. Since third party sales is the main revenue source for the platform, Allegro has every economic incentive to ensure that sellers sell more by helping them to understand their business and make better informed decisions. The table below broadly summarises seller access to the categories of data and analytics provided by Allegro.

### TABLE 4. TYPES OF DATA AND ANALYTICS ACCESSIBLE TO ALLEGRO SELLERS

<table>
<thead>
<tr>
<th>ALLEGRO USER IDENTIFICATION DETAILS</th>
<th>DATA ON TRANSACTIONS BETWEEN BUSINESSES AND CUSTOMERS</th>
<th>BUSINESS PERFORMANCE</th>
<th>USER BEHAVIOUR</th>
<th>ANALYSES OF MARKET TRENDS/DEVELOPMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Own</td>
<td>Competitor</td>
<td>Own</td>
<td>Competitor</td>
<td>Own</td>
</tr>
<tr>
<td>Raw data</td>
<td>Some</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Analytics</td>
<td>NA</td>
<td>Some</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Insights/Actionable guidelines</td>
<td>NA</td>
<td>Some</td>
<td>Yes</td>
<td>Some</td>
</tr>
</tbody>
</table>

#### Data provided to other entities

Allegro also cooperates with public authorities (governments and law enforcement agencies), mostly at individual case level without the provision of bulk data. This can involve sharing of personal data for the purposes of fighting fraud and abuse on the platform or in connection with pending investigations in violation of laws. As is the case with the other larger interviewed marketplaces, the company has a special team that deals with requests from public authorities. More proactive requests from authorities that relate to reaching specific goals or addressing specific challenges, are dealt with in a constructive manner. In certain cases, anonymised data is shared with researchers (e.g. on consumer behaviour or trends) as well as other third-party service providers and trusted partners “to better understand the attractiveness of advertisements and services to Users, improve the overall quality and efficiency of services provided by Allegro.pl or the aforementioned entities”.

Similar to the other interviewed marketplaces, Allegro is aware of the web scraping activities on their website. Measures, such as Google CAPTCHA mechanism, are in place to prevent abusive bot traffic. Moreover, while the primary intended beneficiaries of Allegro’s API are business users and developers, third-party data providers (OMOTs) also download information through APIs as it’s “publicly available”. They can then collate, aggregate and sell this data to interested parties.

The platform also noted to be sharing data based on academic requests. When the data is shared with researchers, it is anonymised and usually focuses on trends and consumer behaviour.

---

AliExpress

AliExpress is a third-party e-commerce marketplace, owned by the Alibaba Group, a Chinese multinational technology company. In 2019, the marketplace was opened to sellers overseas, in: Spain, Italy, Turkey and Russia (through the model of virtual malls)\(^\text{225}\). For example, Spanish sellers are able to sell in 19 EU countries\(^\text{226}\). Currently businesses and individuals can sell on the platform, which charges a commission for each transaction. The seller fees are currently lower than Amazon’s, aimed at attracting a larger number of sellers onto the platform\(^\text{227}\). Aside from small businesses joining the European platform, AliExpress has also approached some well-known brands with invitations to sell their products on the marketplace\(^\text{228}\). AliExpress also has a physical store in Spain, a country which is claimed to be the company’s gateway to Europe\(^\text{229}\). The platform claims to be a second shopping app in Spain, by use and download, with more than 40 million visits per month\(^\text{230}\).

According to its Privacy Policy, AliExpress classifies the information that they collect intro three categories: information provided by users, information collected automatically by AliExpress and information received from third parties. We provide a summary of it in the box below.

**BOX 9. DATA ALIEXPRESS COLLECTS**

- **Information provided by buyers and sellers:**
  - Personally identifiable information such as name, address, phone number, email address and in certain case, ID.
  - Purchase and/or payment information such as credit card or similar payment information, as well as billing and delivery address.
  - Information about purchases (goods purchased and their amount).
  - Information about a seller’s business (if applicable) such as corporation registration, business licences, tax or similar information.
  - Details about the goods and products that the seller intends to sell via the marketplace.
  - Additional information, given with user consent through means such as participation in surveys, e.g. marital status, nationality, gender, date of birth, income, user preferences etc.

- **Information collected automatically:**
  - Information generated through correspondence with AliExpress.
  - Details of transactions and communications over the marketplace.
  - Information on browsing and buying activities, some of which may be considered personal under the EEA data protection laws, e.g. IP address, device types, unique device identification numbers, browser type, geolocation, browsing patterns and details on interaction with the marketplace and its offerings. Furthermore, information on user browser software, operating system, software and hardware attributes, page viewed, number of sessions and unique visitors is also collected.

- **Personal information received from third parties:**

---


\(^{230}\) https://sell.AliExpress.com/es/__pc/itQ5SLp.htm?spm=a2g0o.home.1000001.4.45a170e5VQNishD
AliExpress collects information through its first party cookies, such as analytics cookies, personalisation/advertising cookies, marketing cookies and social networking cookies. For instance, Analytics cookies are used to track browsing behaviour (how users interact with and use AliExpress). The marketplace also uses web beacons - tiny graphics files that enable the recognition and tracking of visitors. Web beacons enable the platform to monitor the traffic patterns of users from one page within the marketplace to another. In addition, AliExpress marketplace uses Google Analytics and allows its third-party cookies to collect data. Through Google’s cookies, AliExpress can analyse user browsing behaviour and website traffic.

Aside from marketing and advertising partners (Google, Facebook, Instagram, etc.), AliExpress may also share the collected data, including personal information, with its other partners and service providers and/or affiliates, e.g. analytics services providers (data related to user behaviour), payment service providers, credit risk assessment and risk control service providers, logistics partners, cloud infrastructure providers, and customer service support providers.

### Data provided to AliExpress business users

There is little information available online on what data is provided by AliExpress to its sellers on its European marketplaces. Based on the information gathered through the desk research (mainly articles on data availability to sellers on its Chinese marketplace), AliExpress provides sellers with data through a Data Latitude tool, which has a set of functions or modules, focusing on different sorts of data. In addition, AliExpress offers a set of tools, such as selection tool, keyword tool, product quality diagnostics tool, ranking query tool (allowing to check the keyword ranking of a product), and data reports. We describe the data provided by AliExpress to its business users in the box below.

### BOX 10. DATA PROVIDED TO BUSINESS USERS THROUGH THE DATA LATITUDE TOOL

- Real-time storm gives access to data (updated every 24 hours) on exposure, browsing, visitors, visitor behaviour etc. This data is claimed to help sellers understand the changes in their store traffic in a timely manner, and determine direct effects of product information optimisation and marketing activities.
- Transaction and sales volume data, transaction status, buyer distribution, top products, search term data (top search keywords), number of orders, supply and demand index and other indexes.

---

331 AliExpress (2019). AliExpress.com Privacy Policy. Available at [https://helpAliExpress.com/buyercenter/questionAnswer.htm?spm=a271m.8038972.0.0.7d06d82v3eYRU&isRouter=0&viewKey=1&id=100099018&categoryIds=92054011&lang=ru](https://helpAliExpress.com/buyercenter/questionAnswer.htm?spm=a271m.8038972.0.0.7d06d82v3eYRU&isRouter=0&viewKey=1&id=100099018&categoryIds=92054011&lang=ru)
332 AliExpress (2019). AliExpress.com Privacy Policy. Available at [https://helpAliExpress.com/buyercenter/questionAnswer.htm?spm=a271m.8038972.0.0.7d06d82v3eYRU&isRouter=0&viewKey=1&id=100099018&categoryIds=92054011&lang=ru](https://helpAliExpress.com/buyercenter/questionAnswer.htm?spm=a271m.8038972.0.0.7d06d82v3eYRU&isRouter=0&viewKey=1&id=100099018&categoryIds=92054011&lang=ru)
333 AliExpress (2019). AliExpress.com Privacy Policy. Available at [https://helpAliExpress.com/buyercenter/questionAnswer.htm?spm=a271m.8038972.0.0.7d06d82v3eYRU&isRouter=0&viewKey=1&id=100099018&categoryIds=92054011&lang=ru](https://helpAliExpress.com/buyercenter/questionAnswer.htm?spm=a271m.8038972.0.0.7d06d82v3eYRU&isRouter=0&viewKey=1&id=100099018&categoryIds=92054011&lang=ru)
334 Cifnews (2015). Seven major data that needs daily attention on AliExpress (interview with Jiang Bo, lecturer of AliExpress University); originally in Chinese. Available at [https://www.cifnews.com/article/15877](https://www.cifnews.com/article/15877)
Industry Intelligence module includes weekly, monthly and quarterly data on selected industries, including industry trends and countries. Sellers can see the top selling industries (categories of products with the highest browsing rate) and the most competitive industries (through top industry rankings). Information is also given on which industry products demand is high (through the statistics on the proportion of transaction orders, proportion of visitors, and proportion of browsing). In addition, sellers can check the shop name of the top seller in each industry and visit their shop. The same module also gives insights into the seller’s store operations and customers, e.g. seller’s product flow, popular products, average transaction price.

- Store Analysis provides data on the store purchase rate, customer unit price, page views, length of visits, and traffic sources
- Commodity Analysis focuses on specific items, their exposure, page views and length of stay, visitors, shopping cart, transaction conversion rate, orders, source of single item traffic, product ranking etc.
- Selection Expert includes most searched items, top selling items.
- Search Word Analysis includes data on top search words and soaring search words
- Daily traffic, traffic sources of the store and individual products (e.g. “paid traffic” (through affiliate or pay for performance marketing, natural search traffic, repeat customer traffic or off-site traffic (social media marketing)), number of visitors (at shop and product level), number of page views, click-through and conversion rates.
- Product Selection Expert function provides merchants with AliExpress’s best-selling products, popular keyword data and industry trends, which is essential for product selection, optimisation and pricing. Keywords can be checked by category and time range.
- Business Analysis function offers specific indicators on the seller’s shop linked to ranking and operation of the shop, e.g. distribution of wireless and PC transactions.

Source: compiled by PPMI, based on desk research

The data provided by AliExpress is said to allow sellers to analyse which products have a high demand, which products are searched for by buyers, and which categories of products have a small number of “blue ocean products” (i.e. products with superior features, untainted by competition, for which demand is to be captured). It is also handy for sellers in terms of developing new products, optimising existing product offerings, and improving the operation of their stores on the marketplace.

We broadly summarise what data and analytics AliExpress makes accessible to its merchants in the table below.

---


## TABLE 5. TYPES OF DATA AND ANALYTICS ACCESSIBLE BY ALIEXPRESS SELLERS

<table>
<thead>
<tr>
<th>ALIEXPRESS</th>
<th>USER IDENTIFICATION DETAILS</th>
<th>DATA ON INDIVIDUAL TRANSACTIONS BETWEEN BUSINESSES AND CUSTOMERS</th>
<th>BUSINESS PERFORMANCE</th>
<th>USER BEHAVIOUR</th>
<th>ANALYSES OF MARKET TRENDS/DEVELOPMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Own</td>
<td>Competitor</td>
<td>Own</td>
<td>Competitor</td>
<td>Own</td>
</tr>
<tr>
<td>Raw data</td>
<td>Some</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Analytics</td>
<td>NA</td>
<td>Some</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Insights/Actionable guidelines</td>
<td>NA</td>
<td>Some</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

### Data provided to other entities

The available information on Aliexpress’ data sharing or access to other third-party entities is very scarce. Nonetheless, one of the main concerns of European consumers and businesses is that, as noted by one interviewee, China gets a “free pass on user data” as under China’s laws, all the data from Chinese platforms goes to the Chinese government. It than can be used by various and unknown purposes (e.g. the infamous Chinese government’s government programme “Social Credit System”, which maybe using data from such ecommerce platforms\(^{237}\)).

### eBay

eBay is a non-vertically integrated marketplace, with an active buyer base of 179 million\(^{238}\). Its main revenue source is hence transaction fees, followed by marketing (classifieds and advertising). As of 2018, eBay was significantly smaller than Amazon in terms of global third-party sales ($95 billion vs. $160 billion)\(^{239}\). Of the $95 billion gross merchandise volume, around 60% was generated outside the U.S., with key countries outside the U.S. being United Kingdom, Germany, Australia, Korea and China.

As the business model of eBay is simpler compared to Amazon’s (with fewer digital services provided), the total scope of data collection is, therefore, significantly smaller too, limited mostly to the activities on its websites and applications. Nonetheless, similar to Amazon, eBay collects a variety of customer and business data when they use the platform (search, transact, etc.), create an eBay account, list a product, provide information via a web form, add or update information in the eBay account, participate in online community discussions or otherwise interact with the platform\(^{240}\). We summarised the data collected and processed by eBay in the box below.

---

\(^{237}\) Cetkovic, N., (2019), “Why China is starting to dominate eCommerce”, IMRG. Available at: [https://www.imrg.org/blog/ecommerce-in-china/](https://www.imrg.org/blog/ecommerce-in-china/)


\(^{239}\) Amazon. (2018). Annual report. Available at [https://ir.aboutamazon.com/static-files/0f9c3eb1-7e1e-4b52-be17-145dc9d8b5ec](https://ir.aboutamazon.com/static-files/0f9c3eb1-7e1e-4b52-be17-145dc9d8b5ec)

**BOX 11. DATA EBAY COLLECTS**

- **Information actively provided by buyers and sellers:**
  - Identification information, such as name, address, telephone numbers or email addresses, user name, VAT identification number, additional identification data and tax identification numbers.
  - Data regarding bids, purchases, or sales that is provided in a transaction.
  - Data provided through eBay messaging tools.
  - Financial information relating to a transaction (e.g. credit card and account numbers, payment details).
  - Shipping, billing and other information related to purchases or shipping of an item, and information required for customs clearance (e.g. tax identification number).
  - Demographic and similar profile information: age, gender, country of birth, nationality, employment status, family status, interests and preferences.

- **Information collected automatically through the use of eBay services, cookies and similar technologies:**
  - Information on bids, purchases and sales that the user is involved in, e.g. transaction amounts, time and location of transactions.
  - Information on actions taken on the marketplace, e.g. placing of an item in the shopping cart, saving sellers, searches or interests.
  - Location data, including the location data of the user’s mobile device.
  - Computer and connection information, e.g. statistics regarding the use of eBay services, information on data traffic to and from websites, referral URL, information on advertisements, IP address, access times, browsing history data, language settings and weblog information.
  - Data on page visits, access time, frequency and duration of visits, links clicked on and other actions.
  - Data about user’s activities and interactions with eBay advertising partners, such as advertisements shown (including how often they were shown, when and where they were shown), whether any action has been taken in response to advertisements, e.g. clicking on an ad or making a purchase.
  - User segment
  - Device information: model or device type, operating system and version, browser type and settings, device ID or individual device identifier, advertisement ID, individual device token, and cookie-related data (e.g. cookie ID), IP address from which the user’s device accesses eBay services.

In addition to this, eBay also collects personal data from other sources, including:

- public sources (e.g. demographic data);
- social media (e.g., data on viewing of videos, ad views and clicks);
- credit agencies or bureaus (e.g. credit reports/checks);
- data providers (e.g. demographic, interest-based and online advertising related data).

---

In data analysis, eBay deals with structured, unstructured, and semi-structured data, where Hadoop, as a big data platform, provides key technology features\(^{243}\). The company is also said to be actively contributing its knowledge back to the open source community\(^{244}\).

Aside from data sharing with consumers and business users (which we cover in more detail in the following section), eBay also transmits data to several other types of entities, although not for re-use, but rather service provision to eBay\(^{245}\):

- Third-party companies that provide services to eBay, such as shipping and carrier providers, payment providers (e.g. operating systems), traffic and marketing partners, and seller partners (e.g. seller aggregators).
- Service providers employed to tackle fraud, to validate customer identity or reduce fraud on the marketplace.
- Government agencies, law enforcement agencies, other national and supranational public authorities or third parties in instances when eBay is asked to participate in proceedings, including judicial proceedings that are conducted for the purposes of tackling, investigating and prosecuting illegal acts.
- Credit agencies bureaus or associations when required by applicable law.

**Data provided to eBay business users**

eBay provides a number of analytics metrics to its business users through eBay Seller Hub or an API connection. Prior to eBay’s acquisition and integration of Terapeak, a previously third-party analytics tool, into its website in recent years\(^{246}\), sellers were provided data through the Listing Analytics app, which was claimed to be the most popular eBay Apps Centre application of all time by the eBay seller community. Sellers reported it to be free, simple to use and generally useful\(^{247}\).

When the acquisition of Terapeak was announced by the platform, it was reported that it underscored the company’s current outlook on building out better data tools for third-party sellers on its platform, to help it compete better with the likes of Amazon and other marketplaces. Basically, the extended data tools were expected to allow the sellers improve and scale their businesses, as well as attract more inventory and sellers to the platform, and therefore increase the general volumes of sales made on eBay\(^{248}\). Indeed, Terapeak seems to offer an expanded package of what was available under the Listings Analytics. The

---


data is said to be more complete, there is a broad range of metrics calculated, it focuses on longer time periods and the search is more specific249.

The comparison between data analytics as was provided under Listings Analytics and as is currently provided under Terapeak tool is summarised in the box below. The Listings Analytics data was available on individual listings, as well as overall averages, for free. Terapeak analytics provides data on sales, trends, velocity, price trends through the sellers’ tools, available through their accounts, i.e. the Seller Hub. Sellers with a Basic, Premium, Anchor and Enterprise eBay Store can access Terapeak for free. All the other sellers using eBay (Starter account holders and non-Store sellers) have the option to pay for a yearly or monthly subscription250. More Terapeak features are being rolled out to provide sellers with additional insights into their business data and that of their competitors251.

**BOX 12. DATA ANALYTICS EBAY PROVIDES**

<table>
<thead>
<tr>
<th>Key data provided under Listings Analytics252</th>
<th>Key data provided under Terapeak Research254</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Sales history, changes in sales, overall performance over time.</td>
<td>• The number of listings and items sold for particular items</td>
</tr>
<tr>
<td>• Selling costs, including specific eBay fees, shipping labels, PayPal fees.</td>
<td>• Average sales prices</td>
</tr>
<tr>
<td>• Competitor’s pricing (also in comparison to the seller’s pricing) and some competitor performance data, also in comparison to seller’s performance.</td>
<td>• The conditions of items sold</td>
</tr>
<tr>
<td>• Listing data, including ranks, click-through rates, page views, sales conversion rates, transactions, where they appear in search results, and how many people see them in results253.</td>
<td>• Sell-through rates</td>
</tr>
</tbody>
</table>

Terapeak hence can be used to evaluate pricing, source opportunities, market trends, and product behaviour across regions and competitors. Among other things, Terapeak also allows sellers to compare themselves to top-performing competitors when it comes to listings, market share average item prices. Upon buyer making an offer on an item, sellers also get consumer data (including email address). While sellers are expected to contact buyers through eBay only, they can also opt for direct communication.

Interviewed eBay seller expressed satisfaction with the quality and quantity of data provided by the platform although some effort is required to actually understand the data. The only issue was the lack of access to historic (i.e., more than three months old) data on exact sales. However, with the introduction of

---


Terapeak, this issue seems to have been addressed as Terapeak enables users to search time periods of any length, across the last 365 days\(^\text{255}\).

The table below broadly summarises seller access to the categories of data and analytics provided by eBay.

**TABLE 6. TYPES OF DATA AND ANALYTICS ACCESSIBLE BY EBAY SELLERS**

<table>
<thead>
<tr>
<th>EBAY USER IDENTIFICATION DETAILS</th>
<th>DATA ON INDIVIDUAL TRANSACTIONS BETWEEN BUSINESSES AND CUSTOMERS</th>
<th>BUSINESS PERFORMANCE</th>
<th>USER BEHAVIOUR</th>
<th>ANALYSES OF MARKET TRENDS/ DEVELOPMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw data</td>
<td>Own</td>
<td>Competitor</td>
<td>Own</td>
<td>Competitor</td>
</tr>
<tr>
<td>Analytics</td>
<td>NA</td>
<td>Some</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Insights/ Actionable guidelines</td>
<td>NA</td>
<td>Some</td>
<td>Yes</td>
<td>Some</td>
</tr>
</tbody>
</table>

**Data provided to other entities**

Generally, data for re-use is not shared by eBay in instances when an external company wants data collected by eBay for own purposes outside of processing of a transaction mandated by the platform. Only on a case-by-case basis, data is provided to researchers based on academic requests\(^\text{256}\). Data provided is not personal but rather focused on trends or user behaviour. Some of the data, such as bid history, is also publicly available.

While web-scraping is not allowed by the platform, and the bots are blocked when identified, it is still occurring. As also seconded by other interviewed platforms, web scraping is increasingly harder to detect because of its sophistication (it is made to look as if it is a normal user browsing the site). The platform interviewee also mentioned that it is possible that competitors are also engaged in web-scraping of eBay’s website.

**Etsy**

Originating from the US, Etsy is a global non-vertically integrated marketplace, aimed at hobbyists and professional merchants that sell handicrafts and vintage items. According to a company representative, around 80% of sellers/ businesses on the platform are of one person and for about 30% of all sellers, it is a full-time job. European markets are important to Etsy: there are roughly 300 thousand merchants active in Europe, representing around 10%-15% of all sellers. Germany, France and United Kingdom are among the six core geographic markets\(^\text{257}\).

---


\(^{256}\) For an example of academic research using eBay datasets, see https://pubs.aeaweb.org/doi/pdf/10.1257/00028280260344632

Intermediation service is the main revenue source of the marketplace (listing, transaction and payment processing fees). Other services, provided by Etsy include the Pattern tool (for personalised website creation), shipping support and ad service.

According to interview with sellers, Etsy collects data on all sellers’ activities on the platform: number and frequency of Etsy shop visits, page views, item views, items sold, number of clicks on the item, response time to a customer, shopping etc. Like the other case study platforms, Etsy obtains data, including personal information, via a few different channels:

— Actively provided by user (registration, form completion etc.).
— Indirectly generated by user activity and interaction with Etsy services, or from the device or browser used to access Etsy services.
— From Etsy vendors and suppliers that the user interacts with (e.g. payment and customer support providers).
— From third-party advertising and marketing partners.

Based on Etsy’s Privacy and Cookies & Similar Technologies Policies, we summarised the data collected or received by Etsy in the box below.

**BOX 13. DATA ETSY COLLECTS OR RECEIVES**

- Information provided by users when registering, setting up an account or using Etsy services, including personally identifiable information such as name, address, telephone number, email address, payment information (credit card information). If a seller decides to use Etsy payment service (“Etsy Payments”), full name, social security number, identification ID or tax ID, date of birth, bank account information, photograph and credit card information is also collected.
- Profile information, including personal information such as birthday, gender, location. Information related to items reviewed, purchased or sold, listed for sale; ratings; reviews made by the user.
- Automated information received and recorded from user’s browser or mobile device when Etsy services are accessed, e.g. IP address or unique device identifier, cookies, page visits, interaction with pages visited, other actions taken on the device.
- Device-specific information, e.g. the hardware model, operating system information, app version, app usage and debugging information, browser information, IP address, and device identifiers.
- Data from Etsy vendors and suppliers, e.g. customer service interactions, payments information, shipping information.
- Data from advertising and marketing partners on user and user behaviour, e.g. attribution information via cookies about the user, UTM tags in URLs (to determine where a visit to Etsy comes from), responses to marketing emails and ads, responses to offers, and audience information.
- Location information: geolocation, IP address, browser information, device information.
- Analytics on user behaviour and app/site performance: how often Etsy apps are used, what happens within the apps, aggregated usage, performance data.
- Information collected via Etsy and third-party cookies and other technologies on the activities on the platform, e.g. account and user preferences, ad targeting, attribution tracking, page views, browsing and purchase activity, demographics, search keywords, favourites etc.

---

Aside from business users, Etsy shares aggregated and de-personalised demographic information with its business partners. Service providers to Etsy (e.g., research companies, and analytics and security providers) are provided with limited access to information but can only use this data to complete tasks on Etsy’s behalf (i.e., cannot re-use the data) and cannot disclose this data to other parties. Etsy also works with a variety of affiliated businesses, which help them perform and improve their services. Some data is shared with them, e.g. in cases when an affiliated business assists in facilitating user transaction. Advertising partners, such as Facebook or Google, are provided with data for advertising purposes, based on user consent. Data is provided in compliance to Etsy’s Terms of Use.

Data provided to Etsy business users

Etsy offers two account types: Standard and Plus. The Plus account holders get access to an expanded set of tools that help them sell but data access does not differ by account type. All sellers see exactly the same information through their shop manager. Etsy shares data through a shop dashboard and via e-mails. Sellers may also download their data in .CSV format and all the data (“shop”) can be downloaded in JSON format. Sellers get access to:

- Transactional data: who bought their product, for how much, where to send etc.
- Aggregated data on business performance through the seller dashboard: sales info, traffic, sources of traffic (e.g. Pinterest, Facebook, Instagram or Google), how many people saw a product or products (views), the connection between traffic sources and sales/views, keywords used to find listings etc.
- Overall business trends and insights, e.g. “purple sells well in dresses”.

These ready-made analytics help sellers to understand what is relevant, what is looked for the most, where the demand is highest in terms of colours, designs, how consumers see the brand, how they order products, and how to improve their traffic. One platform business user noted that statistics on Etsy change all the time (e.g., country views are no longer provided) and it is updated roughly every two to four hours. Etsy ads users also get additional information, related to the performance of their ads on Etsy or Google.

The platform does not provide individual data about the consumers (besides what is necessary for shipment) or other sellers. However, as a result of entering into a transaction with a buyer (when an order is placed), a seller may receive a buyer’s email address or other information, such as name or shipping address. Interviews with sellers also revealed that it is sometimes possible to find customer email addresses of their clients but this information cannot be used outside the scope of Etsy-related communications or transactions without the buyer’s consent.

In addition, sellers on Etsy can use Facebook or Google integration, which means they can get data provided by Google Analytics or Facebook (Commerce manager). Using Google Analytics, sellers can check customer profiles: age, gender, earnings, location, interests etc., which helps to identify target groups and analyse market trends. Since sellers can easily share their listings on Facebook, Etsy has concerns over “disintermediation” by Facebook.

Etsy also offers a Developers’ account, which gives access to much more data through Etsy’s open API. Sellers can register as developers or grant access to third parties (give consent to data use). Data analytics generated through the use of Etsy API can then be sold. Sellers registered as developers can hence offer two services: sell products and data analytics. Through the Developer’s account, one can see how much

---

is sold and for how much (aggregated sales volumes), in what countries, top 10 buyer countries, which companies sell more of a similar product, and so on.

The table below broadly summarises seller access to the categories of data and analytics provided by Etsy.

### TABLE 7. TYPES OF DATA AND ANALYTICS ACCESSIBLE BY ETSY SELLERS

<table>
<thead>
<tr>
<th>ETSY USER IDENTIFICATION DETAILS</th>
<th>DATA ON INDIVIDUAL TRANSACTIONS BETWEEN BUSINESSES AND CUSTOMERS</th>
<th>BUSINESS PERFORMANCE</th>
<th>USER BEHAVIOUR</th>
<th>ANALYSES OF MARKET TRENDS/ DEVELOPMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw data</td>
<td>Own</td>
<td>Competitor</td>
<td>Own</td>
<td>Competitor</td>
</tr>
<tr>
<td>Analytics</td>
<td>NA</td>
<td>Some</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Insights/Actionable guidelines</td>
<td>NA</td>
<td>Some</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Data provided to other entities/ open for collection**

Aside from its business users, service providers and affiliated business, Etsy does not seem to actively provide data to any other third parties for re-use purposes. Nonetheless, like the other marketplaces, it is obliged to provide personal information (such as member’s name, address, phone number, email address, and company name) to public authorities upon lawful requests, to comply with legal proceedings and court orders.

However, Etsy’s data is accessible through its API: access to data can be granted by sellers to third parties. While users of Etsy’s API are required to follow Etsy’s API Terms of Use, including having appropriate privacy and security controls, Etsy does not have a direct partner relationship with these third parties using the API and their processing of information. Reasonable commercial uses of the API are permitted, such as charging users a fee for parts of the application that do not integrate with Etsy API (e.g. advice, consulting, or engaging in the sale of own products or services not competitive with Etsy) or displaying advertising in the developed application. Unless authorised by Etsy, the use of automated systems to access, analyse, or scrape Etsy’s website, including its API, is forbidden. Optimisation tool providers for Etsy sellers, such as Marmalead.com or eRank use Etsy API to access information that Etsy makes available for developers. eRank, for example, has been reviewed by Etsy and can be found in the Etsy App gallery.

**Real.de**

Real.de is a vertically integrated marketplace, with global first party net sales of $182.2 million in 2019. Real.de operates only in Germany, which is their focal market: the website is available in German, the contracting is done in German and the German law applies. Nonetheless, the platform is accessible to European customers and sellers from other countries. Another interesting feature of Real.de is that traders

---

163 https://erank.com/about
on the platform can sell on three other European marketplaces at the same time, through the International Marketplace Network (IMN)\textsuperscript{264}: Cdiscount (France), eMAG (Romania) and ePrice (Italy). Real was originally a supermarket chain (currently running around 280 stationary shops), which decided to expand into online commerce.

The company representative estimated that their e-commerce market share in Germany is around 2-3\%. Unlike Amazon or eBay, also popular in Germany, natural persons cannot sell on Real.de: sellers are only individuals and companies who run a registered business. According to a company representative, there were around 5,000 business users registered on the marketplace as of November 2019, which is a very important and significant part of Real.de business model.

Alongside data actively provided by users of the marketplace, Real.de also collects data automatically and uses first and third-party cookies. Functional cookies (e.g. session cookies) as well as analytics cookies are used to collect user data, such as surfing behaviour. We summarised Real.de’s data collection in the box below.

\textbf{BOX 14. DATA REAL.DE COLLECTS}

\begin{itemize}
\item Data provided by users upon registering and contacting the marketplace, including personal information:
  \begin{itemize}
  \item Name, surname, phone number, e-mail address, gender, billing address, transaction data, date of birth.
  \end{itemize}
\item Data collected automatically on user activities on Real.de website:
  \begin{itemize}
  \item Information about the computer system used to access Real.de website: browser type and version, operating system, IP address, date and time of access, from which website Real.de website was reached, country and place of access. This data is stored for a short time and separately from personal data of users.
  \item Location (GPS) data collected when Real’s mobile apps are used.
  \end{itemize}
\item Data collected via first and third-party cookies:
  \begin{itemize}
  \item Surfing behaviour, interactions with website elements (clicks, mouse position, scroll, hover, blur, focus, anonymised HTML content), product related transaction data, items in a shopping card, language settings etc.
  \item Data on IP address, operating system, browser information, referrer URL (website from which user came from), website and subpages accessed, time spent on the website, data and time of the visit on Real.de website.
  \item Ad conversion tracking: total number of users who clicked on an ad and were redirected to a page with a conversion tracking tag.
  \item Visitor traffic to the website and using behaviour with regard to AdSence ads via Google AdSense.
  \end{itemize}
\end{itemize}

Source: compiled by PPMI, based on Real privacy notice\textsuperscript{265}

Third-party services, such as Google Analytics, Google Optimize, Google Ads, Google AdSense, ContentSquare, Microsoft Bing Ads, Microsoft Azure, Optimizely, Facebook Custom Audiences and App Nexus are used by Real.de for analytics, advertising and marketing purposes. For instance, Real.de uses the third-party services to place ads, track ad conversions, usage behaviour, and visitors’ traffic to and use of Real.de website. These third-parties employ cookies to collect data on Real.de website in order to deliver their services. Similar to the other marketplaces, Real.de can also pass data onto third-parties that

\textsuperscript{264} https://imnnow.com/de/home.html

\textsuperscript{265} Real (n.d.). Privacy. Available at https://www.real.de/rechtliches/datenschutz/#weitere_hinweise_haendler
provide services to Real.de, such as payment service providers, banks, call centres, billing agencies and carriers.

Importantly, the company says to differentiate between marketplace data on the third-party sellers and on own retail business (differently than, for example, Amazon). Due to strict data protection and competition laws in Germany, data from the marketplace (sellers) is not used by Real.de for own business purposes.

**Data provided to Real.de business users**

Little information is available about Real.de data sharing practices. According to a company representative, Real.de provides data that business users need to have to manage their activities and to sell better on the platform. Such data include information on costs, inventory, transactions, sales, and the number of customers. Various data exports, such as information on items offered by a seller and seller orders or product data, can be requested via a seller’s account, under ‘Reports and Statistics’ tab.

In accordance with the German law, Real.de does not provide any comparative analytics about the other merchants on the marketplace or on the overview of other users’ activities on the platform. However, merchants using Real.de advertising services can see information on click rates and conversion rates of different products, but this data is shared in an aggregate form so that merchants can never attribute a product to a certain merchant. The data is provided for free to all merchants active on the platform.

The table below broadly summarises seller access to the categories of data and analytics provided by Real.de. What concerns the other users of Real.de data, the evidence collected did not allow to identify any such actors. When asked about web scraping activities on Real.de website, the company representative said that these activities are obviously happening, but only the data that is available to everyone (i.e. what’s on the frontend) can be scraped.

**TABLE 8. TYPES OF DATA AND ANALYTICS ACCESSIBLE BY REAL.DE SELLERS**

<table>
<thead>
<tr>
<th>REAL.DE</th>
<th>USER IDENTIFICATION DETAILS</th>
<th>DATA ON INDIVIDUAL TRANSACTIONS BETWEEN BUSINESSES AND CUSTOMERS</th>
<th>BUSINESS PERFORMANCE</th>
<th>USER BEHAVIOUR</th>
<th>ANALYSES OF MARKET TRENDS/ DEVELOPMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Own</td>
<td>Competitor</td>
<td>Own</td>
<td>Competitor</td>
</tr>
<tr>
<td>Raw data</td>
<td>Some</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Analytics</td>
<td>NA</td>
<td>Some</td>
<td>No</td>
<td>Some</td>
<td>No</td>
</tr>
<tr>
<td>Insights/ Actionable guidelines</td>
<td>NA</td>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Rakuten**

Often referred to as “the Amazon of Japan”, Rakuten is one of Japan’s largest e-commerce sites, with estimated 365 million monthly site visits per month. Rakuten marketplace is part of Rakuten Group, which is an ecosystem of over 70 cross-sector businesses (online and offline) active in 30 countries and regions. However, unlike Amazon, it is not vertically integrated in the sense that it does not operate as a retailer on its own marketplace. While both, Amazon and Rakuten, are Japan’s leading e-commerce platforms, it is hard to estimate which one accounts for a higher market share since neither of the companies share their total annual transactions. Rakuten quotes its e-commerce market share at 26.8% and says to be dominating the online market in Japan.

Since 2005, Rakuten has been expanding outside Japan. During the time of preparing this analysis, its e-commerce services were active in France and Germany, while marketplaces in the UK and Spain had been closed and only ancillary marketing took place in other EU countries. Rakuten entered France in 2010 by acquiring a leading e-commerce marketplace, PriceMinister. France is the key Rakuten’s market in the EU, but the ecommerce landscape in the country, according to Rakuten’s representative, is extremely competitive. While Rakuten’s European presence is relatively small, Rakuten claims to be the third marketplace in France. Rakuten France accounts for over 22 million members with 10 million active users. Both individuals (buyers-resellers) and businesses/professional sellers can sell on Rakuten.fr. Furthermore, in 2011, Rakuten acquired Tradoria in Germany and rebranded it into Rakuten.de Shopping, which has over 7,000 merchants.

Rakuten actively uses data it collects to further develop the businesses of its ecosystem. The company refers to its shared knowledge and exchange of experience between the Group companies as its advantage over competitors. However, according to an interviewee, in reality, very little data sharing between the teams within Rakuten and among different Rakuten’s services takes place. That is primarily because the data is acquired separately, and it is very difficult to integrate it without notable and continuous investment.

In terms of Rakuten’s data collection through its marketplaces, it is very similar to what the other platforms collect (summarised it in the box below.). Data is provided by users themselves as they create an account, browse the platform or interact with the platform in other ways. Data is also collected automatically by Rakuten, e.g. through cookies or other technologies. Rakuten also allows third-party/partner cookies to be placed on their platforms, such as cookies by Google Analytics or advertising cookies by Google or Facebook. Rakuten website collects and stores consumer behaviour data from which usage profiles are created and then used to analyse visitor behaviour and improve and tailor Rakuten and Rakuten Inc. services.

---


267 https://global.rakuten.com/corp/about/

268 Ibid.

269 https://www.rakuten.co.jp/ec/sellinjapan/

270 https://Infos.rakuten.de/en/about-rakuten


273 https://Infos.rakuten.de/en/about-rakuten


BOX 15. DATA RAKUTEN COLLECTS

- Personally identifiable data of buyers and sellers, such as name, surname, email, postal address, telephone number, date of birth, payment information (payment account name, payment account number, and other payment card details). When creating a professional seller account, Rakuten also asks to identify a legal status of business (e.g. sole proprietorship, limited liability company), registration number of the company and VAT number when applicable.
- Additional information in connection to the user’s account, such as gender or age.
- Device information, such as unique device identifier, device or session IDs, information regarding user’s network and connected hardware, system configuration information, and IP address.
- Browsing, shopping and purchase history.
- Geo-location data as well as date and time stamps associated with transactions (if allowed by user device configurations).
- Navigation data (mostly collected through cookies): this includes anything from the operating system, browser, location data, and interactions with Rakuten content to advertisements on which users click. E.g. search history, the web page visited by the user before he/she comes to Rakuten website, length of visit and number of page views, click-stream data (page-by-page paths taken by the user as he/she browses through Rakuten services, links clicked and conversion information, communications with other members.
- Commercial information, such as records of orders, products or services considered or purchased; information relating to means of payments.
- Inferences about user preferences, characteristics, behaviour and attitudes.

Source: compiled by PPMI, based on Rakuten France Terms of Use and Privacy and cookies policy

Rakuten operated under binding corporate rules on the transfer of data, which are based on the data protection laws of the European Economic Area (EEA). An important background to Rakuten case study, is the EU-Japan free trade deal and especially, the European Commission’s adequacy decision on Japan, acknowledging Japan’s data protection system as adequate. This allows personal data to flow freely between the two economies. In reality, however, it is argued that Japanese practice of data protection and its enforcement is lighter than the EU’s.

Aside from data provision to the business users, which we cover in the following section, Rakuten provides some of the account and browsing behaviour data to their subcontractors, i.e. third-parties offering services to Rakuten. These include payment service providers, shipping and delivery providers, customer service (call centre) providers, web analytics service providers (Google Analytics, Adobe Analytics). Some data generated on Rakuten France platform is also shared with Rakuten parent company in Japan.

Data provided to Rakuten business users

Rakuten offers sellers to integrate their systems with the Rakuten marketplace with a help of its API (known as Rakuten Webservices). It can be used for inventory, sales, stock, accounting management, as well as for data access. Through the API, merchants can obtain information on their own and competing products and listings (allowing to make more informed pricing decisions), sales (status list of ordered items, ongoing sales lists), order history, shipping, and financial transactions. The sellers can also query

---

for list of products by searching for specific key words, category etc. and may receive a list of products with information on product title, image, URL, merchant, the best new price of the offer, and the 10 top new and used listings. Additional services such as sales reports and insights are also available.

Rakuten France professional sellers have the option of accessing the buyer’s contact information, such as email address or telephone number, once the buyer places an order on the Rakuten platform. According to the marketplace’s terms and conditions, these contact details can only be used for the purposes of carrying out the order and have to be erased after the order has been executed.

Finally, sellers using Rakuten Ads service get additional access to data on their ads performance, e.g. the number of impressions, reach and frequency metrics.

The table below broadly summarises the available evidence on seller access to the categories of data and analytics provided by Rakuten.

### TABLE 9. TYPES OF DATA AND ANALYTICS ACCESSIBLE BY RAKUTEN SELLERS

<table>
<thead>
<tr>
<th>RAKUTEN USER IDENTIFICATION DETAILS</th>
<th>DATA ON INDIVIDUAL TRANSACTIONS BETWEEN BUSINESSES AND CUSTOMERS</th>
<th>BUSINESS PERFORMANCE</th>
<th>USER BEHAVIOUR</th>
<th>ANALYSES OF MARKET TRENDS/DEVELOPMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Own</td>
<td>Competitor</td>
<td>Own</td>
<td>Competitor</td>
<td>Own</td>
</tr>
<tr>
<td>Raw data</td>
<td>Some</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Analytics</td>
<td>NA</td>
<td>Yes</td>
<td>No</td>
<td>Some</td>
</tr>
<tr>
<td>Insights/Actionable guidelines</td>
<td>NA</td>
<td>Unknown</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

**Data provided to other entities/open for collection**

Since 2010, Rakuten has been providing some of its data to academic and public research institutions under its Rakuten Data Release initiative. Regarding Rakuten Ichiba (marketplace in Japan), all product data, review data and shop review data has been released. Rakuten France has also opened access to datasets on user reviews, products reviews and book item information. In order to download this data, interested parties need to apply through one of the two Japanese institutions: National Institute of Informatics or the Advanced Language Information Fusion Forum (ALAGIN). Rakuten claims that its open data mission is to 1) contribute to the development of applied technology in academic fields; 2) accelerate evolution cycles in technology by strengthening links between enterprises and academia; and 3) promote unique research by holding symposiums and support application development using large scale data.

---

The data collected did not allow to identify specific companies building their products on data collected on or acquired from Rakuten. However, web scraping activities on Rakuten websites seem to be widespread, given the volume of various tips and tricks materials for that available online. Rakuten, in turn, seems to be fighting this by various safeguards (like CAPTCHA) aimed at identifying robots.

Zalando

Zalando is over 20 years old European online fashion store operating in 17 EU Member States, with 31 million active customers and a turnover of EUR 6.5 billion as of 2019\(^\text{287}\). Zalando has a vertically integrated or hybrid business model and operates both as a traditional retailer and as a marketplace through its Partner Programme. As a traditional retail business, Zalando buys merchandise from brands and resells them for their own price margin to consumers, while the Partner Programme enables brands and retailers to sell their merchandise via Zalando directly to consumers. In terms of revenue split, the Partner Programme is much less significant as it generates roughly 5-10% of platforms’ revenue. Nevertheless, the company is looking to scale their third-party programme to up to 40% by 2023.

There are around 350 brands and retailers selling on Zalando marketplace. A large number of them have a two-fold relationship with Zalando meaning they operate both as sellers on the marketplace and as suppliers to the Zalando wholesale business. Brands and retailers are free to choose under which business model they want to join the platform. Despite its hybrid business model, Zalando claims not to be in direct competition with third-party sellers, e.g. if Zalando and the merchants sell the same item, they offer it different sizes (i.e. Zalando sells S and M directly, whereas L size is sold by the third-party sellers).

As highlighted by a company representative, data collection is crucial for driving the company’s business and they invest a lot into it. An added value of the investment in data collection is a “better fashion journey” for the consumer. Zalando collects different types of data by various means, which we summarised in the box below.

BOX 16. DATA ZALANDO COLLECTS

- Data voluntarily provided by user, including personal and demographic data such as name, address, email address, age, gender, place of residence, various preferences regarding products and brands, device used by the user, content of communication with Zalando. Other information shared by users, such as photos or other personal content.
- Purchasing data (order number, details about purchased items, payment method details etc.).
- Payment details (preferred method of payment, bank account details, credit data etc.).
- Data generated by interaction of the user with Zalando services, such as interest data in terms of product types, brands and styles. This also includes data derived by Zalando based on usage analysis of browsing behaviour.
- Data from third parties, such as credit rating agencies, payment services providers as well as advertising partners (in this case, only aggregated, encrypted or anonymous data is received on demographics, device and access data, and interests).
- Social network public profile data, in case where Zalando’s services are integrated with social networks, such as Facebook.

Zalando uses the collected data for various purposes, primarily for purchase processing, and the provision, personalisation and further development of their services. This also includes using data for data analysis, advertising and marketing research purposes, such as user segmentation, insights into the shopping and usage habits, purchasing behaviour, detection of fashion trends. Zalando also shares data with external service providers e.g. hotline service providers and logistics companies that need it to provide services to Zalando. These service providers can access and process data only to the extent and for the period that is necessary for the provision of services to Zalando. Similar to other marketplaces, the company also passes data to its advertising partners, such as Google and Facebook for the use of their services, mostly advertising. The data shared concerns target groups for certain merchandise.

Since the company considers its two business channels (retail and marketplace) as competing, different data sharing arrangements are in place for Zalando’s suppliers and the third-party sellers. Generally, for suppliers to Zalando wholesale, aggregated data is provided via the CoPlanner tool while third-party sellers on Zalando marketplace get access to data via zDirect. We cover data provision to third party sellers in the following section.

Data provided to Zalando business users

Zalando provides a variety of data to its third-party sellers via zDirect, which the company considers to be relevant, e.g. number of units sold, at what price etc. A summary of the data provided by Zalando to its third-party sellers is provided in the box below.

**BOX 17. DATA THAT ZALANDO PROVIDES TO ITS SELLERS**

- On current business: data about the merchandise that partner companies have sold on Zalando marketplace, such as sold items, type of products, when it has been sold, as well as pricing. The brands should already be aware of most of this data because the brands input it by setting the price and quantity, monitoring items sold.
- Customer insights: data that helps brands to understand who their customers are or what customers they are serving. This data is especially useful for smaller brands that do not have their own data analytics. It is quite straightforward for the platform to provide brands with certain information on their customers, such as gender identification, but it becomes trickier with providing other information, e.g. on where the customers are based, which trends they follow etc.
- Customer identification details: sufficient information is provided for carrying out invoicing and sending parcels to the customer (e.g. email address or telephone number). For commercial reasons, Zalando does not allow their partners to directly target customers that are not a part of the marketplace business (i.e. due to the risk of disintermediation).
- Market overview: overview of brands’ respective markets that helps them better plan for the future and understand changing customers’ behaviours. This is especially relevant in the fast fashion sector, which is developing very rapidly and in which brands are expected to adapt very quickly to customer expectations.
- Competitor data: Zalando offers high level aggregated information on competitors, such as what is their standing is, how they can improve within their category. For example, in the sports sector Zalando informs partners how they are doing in comparison to a selected number of other brands. Zalando does not mention

---

companies, just the biggest partners. They do not share more insights on the brands separately (for example, brand X has this volume of revenues), and just insights in aggregate form are shared.

Source: compiled by PPMi, based on information from interviews and desk research.

Sharing data concerning businesses that compete with each other on the platform, is, to some extent, subject to regulation by national law. In Germany, sharing information on revenue from a specific product to the competing brand could be seen in terms of the anti-trust law as providing competitive advice on revenue or pricing. Zalando therefore shares only highly aggregated data on other sellers active on the marketplace.

A company representative also noticed in an interview that that third-party sellers are interested in information that goes beyond the scope of the marketplace. For instance, how the seller’s merchandise is performing within Zalando’s retail business, if it also acts as a supplier. By anti-trust law, however, Zalando can share only specific information on certain number of key performance indicators (KPIs).

In the context of scaling its Partner Programme, Zalando is planning to provide deeper insights for sellers rather than more data. This should be useful for the larger retailers, as the smaller sellers, according to the platform’s representative, do not make use of the data that currently Zalando provides.

The table below broadly summarises the available evidence on seller access to the categories of data and analytics provided by Zalando.

### TABLE 10. TYPES OF DATA AND ANALYTICS ACCESSIBLE BY ZALANDO SELLERS

<table>
<thead>
<tr>
<th>ZALANDO USER IDENTIFICATION DETAILS</th>
<th>DATA ON INDIVIDUAL TRANSACTIONS BETWEEN BUSINESSES AND CUSTOMERS</th>
<th>BUSINESS PERFORMANCE</th>
<th>USER BEHAVIOUR</th>
<th>ANALYSES OF MARKET TRENDS/DEVELOPMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Own</td>
<td>Competitor</td>
<td>Own</td>
<td>Competitor</td>
<td>Own</td>
</tr>
<tr>
<td>Raw data</td>
<td>Some</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Analytics</td>
<td>NA</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Insights/Actionable guidelines</td>
<td>NA</td>
<td>Unknown</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data provided to other entities/ open for collection

The data collection activities did not allow to identify specific types of other business businesses or specific entities gathering and re-using Zalando’s data. However, Zalando does not prohibit web crawling or scraping but has measures in place that make it more difficult for various actors to engage in such activities. This is done due to 1) operational or pragmatic reasons as web crawling slows-down the website, making it slower and therefore less attractive to consumers; 2) security reasons.

Google Play

Google generally collects user data in a variety of ways. The most obvious is the active data provision, with the user directly and consciously communicating information to Google, as for example by signing-
in to any of its widely used applications such as YouTube, Gmail, Search, etc., or developer listing and distributing an application via Google Play. Less obvious ways for Google to collect data are “passive” means, whereby an application is used to gather information while it is running, possibly without the user’s knowledge. Google’s passive data gathering methods are integrated into its platforms (e.g. Android and Chrome), applications (e.g. Search, YouTube, Maps), publisher tools (e.g. Google Analytics, AdSense) and advertiser tools (e.g. AdMob, AdWords). The full scope of data collected is generally difficult to grasp to outsiders; academic studies have been conducted on the issue, and various whistle blowers occasionally bring to attention new ways and areas in which the platform collects user data. Some findings indicate that Google has the ability to connect the anonymous data collected through passive means with the personal information of the user.

Given that Google Play app store is tightly connected to Google’s ecosystem through Android OS, which is the key enabler of data collection for Google, it is difficult to strictly delineate user data that is collected exclusively on Google Play from data generated in Google’s other activities and the very apps listed on the store (e.g. Chrome). Obviously, it collects information on all the activities and transactions going on the app store, as well as transactions within apps. However, an interviewed Google Play’s representative argued that “it’s a common misconception that [the Play Store] collects everything”. On the contrary, their objective is to be very strategic about data collection, “so that it could be stored safely and securely, and be compliant with GDPR”. As explained by Google Play’s representative, as an app store and payment processor, they track:

- user installs and uninstalls,
- purchases and refunds,
- signals of abuse and fraud,
- user settings,
- user generated content, including rankings and reviews,
- detailed user activity on the app store,
- specific device information.

Data provided to app developers

The app store claims that they aim to provide the developers with the optimal amount of data to successfully operate and grow their apps. The platform presents various statistics to its app developers for their individual apps (on the Play Console web version or app; the main categories are detailed in the box below) as well as extensive material on how to use this data in decision-making. Some developers

---

report that this information enabled them to improve customer retention, engagement and monetisation\(^n\).

**BOX 18. GOOGLE PLAY CONSOLE: KEY METRICS FOR APP DEVELOPERS**

**Statistics and insights**: developer can find and review their app statistics on their computer or Android apps. There are several pages within the Play Console where they can review their app’s installs, uninstalls, ratings, revenue, and crashes data. Dashboard page provides overview of key metrics, trends, notifications, and insights; Statistics page provides customizable, detailed reports with key metrics and dimensions. Some metrics are calculated based on data from users who have agreed to share their data with developers in aggregate, but adjusted to more closely reflect data from all of the users of that specific app developer. For example, a developer’s data can be compared to peer groups, over time or before-after the dates of some key events (e.g., subscription price change, roll-out of new releases). The reports can be accessed and downloaded as CSV files from Google Cloud Storage. They are generated daily and accumulated in monthly CSV files. More specifically, the provided metrics cover:

- install-related statistics (e.g., users and devices (new, active, returning), install and uninstall events);
- ratings (e.g., average and cumulative average rating, rating volume, Google Play rating);
- revenue (e.g., total revenue, revenue per time period, buyers);
- crashes and application not responding errors (crashes and ANRs);
- specific metrics for Android Instant Apps (launches by device, launch events, conversion events).

The Play Console Help portal provides support to developers, explaining, for example, how to analyse ratings and reviews, measure apps acquisition and retention, monitor app’s technical performance, compare ratings with custom peer groups, use the data to stop a staged rollout, run A/B tests on store listing and apply the results and manage app’s orders/subscription cancellations - among other things.

In addition to this, developers can use the ‘Growth rate compared to peers’ chart to understand how their metrics are performing in relation to groups of peer apps from one period to the next. Developers can create a custom set of 8-12 peers to compare their app to, then see the median value of the set and the difference between their app and its peers for Android vitals data as well as for public metrics like rating. This data allows developers to judge their performance against key areas of the ecosystem.

**Monetization**: Google Play also provides information on a developer’s app’s revenue and buyer data. They can review app’s financial data to see how their sales, managed products, subscriptions, and rewarded products perform over time. Revenue data is based on estimated sales (amounts paid by buyers, including tax). The data categories include:

- Overview - a high-level breakdown of revenue sources, average revenue per paying user (ARPPU), and buyer information.
- Revenue - per product, individual product details, export revenue data.
- Buyers - total, new and returning, by device and country; using time filters.
- Conversions - rate, spending per buyer.

The app store offers two types of transaction reports:

- Earnings Reports: Generated near the beginning of the month.
- Estimated Sales Reports: Generated daily by adding all transactions that were CHARGED or REFUNDED recently to the current month’s file. It can take several days for all new transactions to appear.

An interviewed representative of a developer association summarised that the data that developers get from platforms do not necessarily show how other businesses are doing, but provide a very good understanding how the developer’s own apps are performing. In addition to this, the platform invests in  

---

developing various materials aimed at providing tips and guidelines for smaller developers to improve their products, adjust their monetisation models and grow.

The Google Play representative explained that the metrics provided are continuously reviewed and updated, with two goals in mind: usefulness to the developers and customer data protection. On the one hand, they gather developer feedback to find out what metrics would be useful to introduce. On the other hand, they may limit the access to protect the end users. For example, since May 2018 some metrics have been calculated based on data from users who have agreed to share their data with developers in aggregate. The metrics are adjusted to more closely reflect data from all of individual developer’s end-users, but Google does not display data that falls under certain minimum thresholds anymore.

Out of the data that it collects, Google Play does not provide developers with user device information, detailed activities on the Play Store or metrics on fraud detection. Although Google Play is a vertically integrated platform (i.e. it distributes Google first-party applications), the platform’s representative assured that according to their strict internal policies, first-party app developers have the same access to Google Play data as the third-party app developers. More specifically, the Play Store restricts first-party developer teams from accessing Google Play data about competing third-party developers (e.g. the Chrome team is not allowed to see competitively sensitive developer data and metrics about a rival web browser app by a third-party developer). Nonetheless, as part of Google, their first-party developer teams may have access to aggregated, non-identifiable data about various parts of the business that is not accessible outside the company.

Furthermore, the data that developers need the most concerns the user engagement with their own apps - something that Google Play says it does not collect, and leaves it up to developers themselves to collect. Therefore, the data that developers have (or do not have) access to depends on the end-user permissions. The app store has certain requirements for app privacy policy, but the scope of data (including personal and sensitive data) that individual apps can access is based on requests made directly to the users and the permission that users are asked or choose to grant on Android system. Android’s permissions system is based on the security principle of least privilege: an app should only have the minimum data needed to perform its task. Developers must declare the permissions that their apps need beforehand, and the user is given an opportunity to review them and decide whether to install the app. Interestingly, however, a recent study by the International Computer Science Institute (ICSI, the US) claims that thousands of Android apps can access restricted data even when users deny them permission. Google said they would address the issues in Android 10, which was released in September 2019.

The broad categories of data and analytics accessible for app developers active on Google Play are summarised in the table below.

---

TABLE 11. TYPES OF DATA AND ANALYTICS ACCESSIBLE BY GOOGLE PLAY APP DEVELOPERS ON THE PLATFORM

<table>
<thead>
<tr>
<th>GOOGLE PLAY</th>
<th>USER IDENTIFICATION DETAILS</th>
<th>DATA ON TRANSACTIONS BETWEEN BUSINESSES AND CUSTOMERS</th>
<th>BUSINESS PERFORMANCE</th>
<th>USER BEHAVIOUR</th>
<th>ANALYSES OF MARKET TRENDS/ DEVELOPMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Own</td>
<td>Competitor</td>
<td>Own</td>
<td>Competitor</td>
</tr>
<tr>
<td>Raw data</td>
<td>Depends on user permissions</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Analytics</td>
<td>NA</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Some</td>
</tr>
<tr>
<td>Insights/ Actionable guidelines</td>
<td>NA</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

Data provided to other entities/ open for collection

The interviewed Play Store representatives explained that they do not really know what other parties are using their app store data. This is because it is set up for the app developers and end-users, and not for any other parties.

They do not make any data available specifically for third-party ASO, analytics or marketing providers. These get app store information from several sources. One is web-scraping. The app store is aware if the data scraping activities take place on the store, but generally the information that could be collected this way is limited. Some of the publicly available data on apps of Google Play is also being used in academic research\(^9\). Secondly, ASO providers receive data from the accounts of developers who provide app store data about themselves voluntarily. According to Play Store’s terms and conditions there are no restrictions for developers to share this data with third party companies.

Apple App Store

Differently from Google, which operates the Play Store, Apple does not engage in behavioral advertising business (although the business models of both app stores are commission fee based). This, of course, influences its incentives to collect and use the user behavior data concerning the activities taking place on the App Store, and on the Apple ecosystem of apps, services and devices more broadly. The company claims that customer privacy is at the core of the company’s DNA. It is part of the company’s success and the guiding the principle for data collection, use and sharing practices.

Based to the company’s privacy statement, Apple collects both personal and non-personal information on the users directly from them or other sources, and combine in various ways to create inferred datasets. More specifically, the company discloses to collect data in the following ways:

— When users create an Apple ID, apply for commercial credit, purchase a product, download a software update, register for a class at an Apple Retail Store, connect to services, contact Apple including by social media or participate in an online survey, Apple may collect a variety of


98
information, including your name, mailing address, phone number, email address, contact preferences, device identifiers, IP address, location information, credit card information and profile information where the contact is via social media.

— When users share their content with family and friends using Apple products, send gift certificates and products, or invite others to participate in Apple services or forums, Apple may collect the information provided about those people such as name, mailing address, email address, and phone number. Apple will use such information to fulfill your requests, provide the relevant product or service, or for anti-fraud purposes.

— In certain jurisdictions, Apple may ask for a government issued ID in limited circumstances including when setting up a wireless account and activating device, for the purpose of extending commercial credit, managing reservations, or as required by law.

— For research and development purposes, Apple may acquire and use datasets such as those that contain images, voices or other data that could be associated with an identifiable person.

— Apple also may collect information such as occupation, language, zip code, area code, unique device identifier, referrer URL, location, and the time zone where an Apple product is used to better understand customer behaviour and improve our products, services, and advertising.

— Apple may collect information regarding customer activities on our website, iCloud services, our iTunes Store, App Store, Mac App Store, App Store for Apple TV and iBooks Stores and from Apple’s other products and services.

— They may also collect and store details of how consumers use services, including search queries.

— With customer’s explicit consent, Apple may collect data about how consumers use their devices and applications in order to help app developers improve their apps (this then feeds into app analytics provided to developers - see the section below).

The company shares the information it collects only to provide or improve their products and services; or if data sharing is mandated by law. Meanwhile the information collected by app developers is governed by their privacy practices, within the applicable legal framework.

Data provided to app developers

App developers are the only intended beneficiary that the platform serves in terms of data sharing. Apple representatives claim that they provide the developers with a robust set of tools and information about how their apps are performing on the platform. The information on app performance is provided at an aggregated yet quite detailed level.

All developers, no matter what their applications or monetization models are, have the same access to these tools. The data provided is a part of one standard package that offered to all app developers on the App Store, and there is no possibility to get additional data from the platform for additional fees, according to the interviewed platform representatives.

Consumers can opt-in to sharing their non-identifiable data with developers. Apple provides the developers the data analytics including where the customers are coming from, the volume of app launches and other statistics derived from use of the app - basically everything on how the consumers got to the app and how they are using it. Consumer consent is sought from users when they launch their devices for the first time or after a major software update, where Apple asks if the customers are willing to share personally non-identifiable information with Apple, explains the reasons why, and separately asks on the same screen if they are willing to share personally non-identifiable information with app developers. The

---

customer decides separately whether they want to provide data analytics to Apple and to third-party developers.

Apple teams that develop first-party apps, according to the platform’s representatives, do not have exclusive access to the data and analytics on third-party App Store developers, some of which may have listings in direct competition with Apple’s first-party apps.

Any personally identifiable user information collected by the platform, according to the App Store representatives, is off limits for the developers. Apple claims to make effort to ensure that app developers cannot identify and trace back to the individual users that use their apps, in order to meet customer expectations. The terms and conditions of developer’s access and use of data require that no attempt is made to re-identify the end-users.

The key metrics on the performance of their applications are provided to developers though the App Store Connect portal, accessible through a web-browser or an online app. The portal is created for developers to manage all aspects of their app businesses on Apple. There they can also access App Analytics and Sales and Trends sections. The key metrics provided there are presented in the boxes below.

App Analytics was first introduced in 2015, and since then it has been continuously updated with new metrics (e.g., during the time of preparing this analytical paper, the “Deletions” metrics were the latest addition to the developer analytics toolkit provided by the app store). Other ways to track similar metrics have been third-party integrations (as described in Section 3.2.2.), sometimes offering more detail. However, Apple’s App Analytics has the advantage of being a first-party measurement solution, ensuring that the data is more reliable. For instance, only Apple can provide the date that a user with an Apple ID first downloaded an app from the App Store and some other important data. Meanwhile, the third-party providers largely rely on inferred insights based on the indicators they can scrape or collect through other means.

### TABLE 12. APPLE APP ANALYTICS METRICS AND DIMENSIONS 302

<table>
<thead>
<tr>
<th>METRICS</th>
<th>METRICS AND DIMENSIONS (2020)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usage</td>
<td>- Active devices</td>
</tr>
<tr>
<td></td>
<td>- Active in last 3 days</td>
</tr>
<tr>
<td></td>
<td>- Crashes</td>
</tr>
<tr>
<td></td>
<td>- Deletions</td>
</tr>
<tr>
<td></td>
<td>- Installations</td>
</tr>
<tr>
<td></td>
<td>- Sessions</td>
</tr>
<tr>
<td>Sales</td>
<td>- App Units</td>
</tr>
<tr>
<td></td>
<td>- In-App Purchases</td>
</tr>
<tr>
<td></td>
<td>- Paying Users</td>
</tr>
<tr>
<td></td>
<td>- Sales</td>
</tr>
<tr>
<td>App store</td>
<td>- Impressions</td>
</tr>
<tr>
<td></td>
<td>- Product Page Views</td>
</tr>
<tr>
<td>DIMENSIONS</td>
<td>- App Purchase Date</td>
</tr>
<tr>
<td></td>
<td>- App Version</td>
</tr>
<tr>
<td></td>
<td>- Campaigns</td>
</tr>
<tr>
<td></td>
<td>- Device</td>
</tr>
<tr>
<td></td>
<td>- Platform version</td>
</tr>
<tr>
<td></td>
<td>- Purchasable item</td>
</tr>
</tbody>
</table>

301 https://developer.apple.com/app-store-connect/analytics/
Each of the metrics presented in the table above can be broken down by each of the dimensions. The platform also provides explanations how to use these metrics to further estimate other indicators important for intelligent decision making. For instance App Store Impressions can be used to see how often users click through to download an app from seeing it anywhere on the App Store; User engagement indicators allow to evaluate the impact of product changes (e.g. modifying the initial onboarding experience); Deletions may also show user reactions to crashes or changes in the app.

Another set of data and analytics helping app developers to measure their app performance is Sales and Trends reports, providing next-day sales and unit data. They include the information listed in the box below.

**BOX 19. APPLE APP SALES AND TRENDS REPORTS**

<table>
<thead>
<tr>
<th>SALES AND TRENDS REPORTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary Sales Report: Aggregated sales and download data for apps and In-App Purchases.</td>
</tr>
<tr>
<td>Sales Events Report: The dates of significant sales events for apps and in-app purchases across various territories and devices.</td>
</tr>
<tr>
<td>Subscription Report: Total number of Active Subscriptions, Subscriptions with Introductory Prices, and Marketing Opt-Ins for your auto-renewable subscriptions.</td>
</tr>
<tr>
<td>Subscription Event Report: Aggregated data about subscriber activity, including upgrades, renewals, and introductory price conversions.</td>
</tr>
<tr>
<td>Subscriber Report: Transaction-level data about subscriber activity using randomly generated Subscriber IDs.</td>
</tr>
<tr>
<td>Opt-In Report: Contact information for customers who opt in to share their contact information.</td>
</tr>
<tr>
<td>Pre-Order Report: Aggregated data for apps made available for pre-order, including the number of units ordered and canceled by customers.</td>
</tr>
</tbody>
</table>

From April 2020, the app store introduced the third section of Insights available to app developers through the App Store Connect portal: Payments and Financial Reports. They provide more details on the final proceeds of apps and in-app purchases, including the transaction and settlement dates for purchases, currencies used, as well as the geographic breakdown of transactions. Other most recent update to data access includes server-to-server notifications providing real-time updates on a subscription’s status, so developers can provide customized experiences for subscribers. According to Apple, these and other features increase the level of data granularity that developers are requesting, with the aim to make their performance metrics more meaningful. If developers have marketing campaigns outside Apple, they can merge Apple data with other datasets from other relevant sources to see how their campaigns are doing.

Generally, the app store representatives claim that Apple’s teams are in constant contact with the developers, where they get feedback on data and analytics that developers would like to see. They are regularly updating tools available on App Store Connect, including data analytics tools.
### TABLE 13. TYPES OF DATA AND ANALYTICS ACCESSIBLE BY APPLE APP STORE DEVELOPERS ON THE PLATFORM

<table>
<thead>
<tr>
<th>Apple App Store</th>
<th>User Identification Details</th>
<th>Data on Transactions Between Businesses and Customers</th>
<th>Business Performance</th>
<th>User Behaviour</th>
<th>Analyses of Market Trends/Developments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Own</td>
<td>Competitor</td>
<td>Own</td>
<td>Competitor</td>
<td>Own</td>
</tr>
<tr>
<td>Raw data</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Analytics</td>
<td>NA</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Insights/Actionable guidelines</td>
<td>NA</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Data provided to other entities/open for collection**

As the representatives of the Apple App Store explained, their systems are built to share the data with their developers. Other entities are not served in such a way.

Speaking about the *app store optimisation tools*, the App Store claims that they do not have a feed of data to such providers, and they do not support such entities in any way. They are generally not aware how the ASO tools get the data concerning app stores. Web scraping is not permitted under the Apple App Store T&C, but there is not much they can do to stop it, and this is not a “hot topic internally” at Apple.

Importantly, as mentioned above, most of the data important to developers they *can collect themselves*. Given that many other platforms operate as apps, they also seek to gather data though their apps listed on the Apple App Store. This, however, is sometimes portrayed in the media as a source of controversy, in which the App Store has not been sufficiently protective of consumer privacy (opposite to the criticism that app stores provide too little data). For example, in 2019 Apple got into a dispute with Facebook, as the latter had been paying people, including teens 13 to 17 years old, to install a “research” app that extracted huge volumes of personal data from their iPhones—direct messages, photos, emails, and more. Facebook used this information partly to improve its data profiles for advertisement, but also as a business-intelligence tool to investigate competitor behaviour. Furthermore, in 2018 Google reportedly paid Apple $9 billion, and $12 billion in 2019 to keep Chrome a default search engine on Apple devices. All those searches help funnel out enormous volumes of data on Apple’s users to the other big platform players.

---


Booking.com

Founded in 1996 in the Netherlands, Booking.com, a hotel booking platform, is one of the major European online platforms, active globally. It is owned and operated by the US-based Booking Holdings, and is headquartered in Amsterdam. Other brands of the group include Kayak, Priceline, Agoda, Rentalcars.com, and OpenTable.

Much like other multi-sided platforms, Booking.com collects data on its users both actively and passively. More specifically:

— Some of the data, the customers (or e.g. travel agents on their behalf) provide directly to the platform. For example, to make a reservation, the minimum required data is name and email address. Usually, however, the platform also asks for home address, telephone number, payment information, date of birth, the names of guests travelling together and preferences. Further, the platform requests to leave reviews for the booked properties after the stay, which many customers do.

— Booking.com also collects customer data from their mobile devices, if the permission is given (e.g. location data, contact details).

— Customer participation in referral programs or sweepstakes provides Booking.com with additional personal data.

— Most of the customer data is collected automatically as they visit and navigate Booking.com and affiliated websites or apps. This includes IP address, the date and time they accessed Booking.com services, the hardware, software, internet browser and computer’s operating system, app version and language used, clicks and pages shown to the customer. From mobile devices, the platform collects device identification data, device-specific settings and characteristics, app crashes and other system activity.

— In addition to this, Booking.com receives customer information from other sources, including “affiliate partners, subsidiaries of the Booking.com corporate family and other affiliates of the Booking Holdings Inc. corporate family and other independent third parties”\(^{305}\). All the information that the platform acquires from these sources may be combined with information provided by customers themselves.

— Booking.com also monitors all communications between customers and service providers that are conducted through Booking.com communication means.

— If Booking.com account is linked to social media accounts, Booking receives data from the social media platform, such as cookies and other data used for advertising\(^{306}\).

Significant amounts of data are also collected on the business users, including detailed and comprehensive information on the listed properties and their policies, pricing, transactions, free and booked rooms and so on. Generally, the platform’s data flows result into a lot of personal information of customers, as well as detailed information on around 28 million properties\(^{307}\) using the Booking.com platform. This, in turn, also provides the platform with insights on destinations and countries all around the world.


According to the platform’s Privacy and Cookie Statement\textsuperscript{308}, the data that Booking.com collects is shared with several types of third parties, with the main aim to successfully conduct its business, deliver its services and complete the transactions:

- “Trip Providers” - the platform’s business users.
- Business partners (e.g., for distribution or advertising).
- The Booking Holdings Inc. corporate family and local Booking.com offices.
- Third-party service providers, including for customer support, market research, fraud detection, payment and marketing services.
- Public authorities, as required by law.

Further we review those third-parties that receive or collect platform data for re-use.

**Data provided to business users**

The platform claims to provide the listed properties with all the tools, including data, to help their business succeed. Booking.com Connectivity APIs allow businesses to send and retrieve data for their properties listed on Booking.com. They can manage room availability, reservations, prices, and many other things within their own systems or outsource these processes to third-party providers by giving them this access. Moreover, they can access the platform’s data on room rates, reservations, promotions, and guest reviews. This enables business users to build a “one-stop shop” for their connected properties, allowing property owners to easily manage their information on multiple websites\textsuperscript{309}, including other OTAs.

Furthermore, Booking.com conducts analytics on these data to provide hotels with data services, such as pricing strategy, demand forecast, and consulting services. Booking.com Partner Hub portal provides an analytics dashboard for the platform’s business users, including some aggregated data on their competition\textsuperscript{310}. According to the platform, the aim is to help hotels boost their sales and find ways of getting more bookings. The box below provides a detailed list of items that Booking.com provides to its business users in the Analytics section.

**BOX 20. ANALYTICS FOR BOOKING.COM BUSINESS USERS**

- Analytics dashboard, offering Business reports with real-time information, helping to monitor past and future sales, as well as compare the user’s hotel’s results against their competitors;
- Booker Insights report, that gives some information on the hotel’s current bookers, helping to better understand who their customers are\textsuperscript{311};
- Information useful for reservations management, for example, how far in advance customers are booking their rooms, the number of room nights sold, sold out dates, as well as revenue broken down in a range of charts.
- Guest review scores’ tool, allowing to analyse user review scores over time and to compare performance against properties in their competitive set\textsuperscript{312} (~10-15 properties).


“Genius” programme reports (for participating properties), including statistics on Genius guests and gains (additional revenues) from the programme.313.

Visibility Dashboard, providing data to track and improve property’s visibility in Booking.com search results, such as:

- search result views (the number of times potential guests saw the property in search results);
- click-through rate (CTR; the percentage of potential guests that clicked on the property in the search results, relative to the number of search results views);
- property page views (the number of times potential guests viewed the property page);
- bookings (the total number of bookings received, including cancellations) and the overall conversion rates.
- visibility performance in Booking.com search results the past 30, 90 or 365 days.
- comparative visibility performance against other properties in the market for the past 90 days (see the figure below).

- Conversion (how well the property converts guest searches into bookings).
- Key factors that influence performance, compared to other properties in the area, and pointing to possible actions to increase exposure to more potential guests (see the figure below).314.

---

Since 2016 when it was rolled out\(^{315}\), the Analytics section has been continuously updated with new features and metrics. According to a Senior Product Manager at Booking.com, they spend a lot of time to offer information that would help the business users to provide “better guest experience and creating something scalable”. To expand the list of metrics and insights provided, they look for a “sweet spot” between what the business users ask for (i.e., what they know they would use and are not getting from customers) and information that guests are willing to share\(^{316}\).

In addition to this, the BookingSuite division of Booking.com provides its partner hotels with additional data solutions, including RateIntelligence - a free tool for smaller properties\(^{337}\), providing market data (demand, room rates, local events) for informed pricing decisions and price optimisation. It was reported by the platform that Booking.com data analytics service on pricing strategy on average increased third-party sellers’ sales revenue by 7\(^{339}\). Business users of Booking.com also reported having access to broader travel trends, such as popular days for travel, what types of properties make more sales (e.g. 3-star hotels vs 5-star hotels vs hostels). Upon request, the platform can provide additional insights to business user account managers at Booking.com (e.g., impact of Brexit on hotel bookings).

Furthermore, the platform also offers its business users a Commercial Insights\(^{320}\) section, containing various tips on travel seasonality and ways to keep their rooms filled. Also, in late 2019, Booking.com launched Guest Insights tool for their business users\(^{331}\). Eligible hotels (i.e., those with guest reviews averaging 8 or more during the previous quarter and replying (within a period of 24 hours) to 70% or more of the guest messages you received during the previous quarter\(^{372}\)) get information on their guest interests, age group, allergies, dietary restrictions and past reviews. This is intended to allow the properties to provide their guests a more personalised experience\(^{323}\). The insights are available for the business users on the platform’s Extranet or the Pulse app for the platform’s partners.

Basic data analytics package as described above is mostly the same for all business users, but bilateral strategic discussions are held with some bigger partners in order to improve their business results on Booking.com. The platform, therefore, may differentiate analytics access by the size of partner to some extent. Large hotels with many rooms listed on the platform get more platform’s attention and more analytics.

Importantly, Booking.com, as many other OTAs does not provide the hotels with full customers contact data (i.e., they share phone number, but not email address). All the communication between properties and their customers goes through Booking.com internal communication tools\(^{334}\). When an OTA booking is made, hotels are not passed along any first-party data (apart from that necessary for the transaction),

\(^{337}\) While larger properties with own IT departments can get this data through an API and analyse it themselves.
leaving the only chance to capture customer contact data during the visit\textsuperscript{325}. The customer contact data, meanwhile, is hotels’ chance to ensure personalised service, as well as build guest relationships through loyalty marketing and promotions\textsuperscript{326}. In fact, some of the interviewed hotels tended to disregard all the data and analytics provided by the platform as unimportant, and underlined that the customer contacts and identification details are the key data that the OTA refuses to provide, causing them various issues.

However, related to the possibility of bilateral discussions described above, the largest players in the hotel industry also have different access to Booking.com customer personal and contact data. Although the standard terms and conditions are the same for all hotels, the largest hotel chains reportedly manage to negotiate access to some more customer data. This does not, however, mean that they have full access to their consumer details after a booking is made. Even the biggest global hotel chains account for a miniscule share of the total listings available on Booking.com, and their negotiating power is therefore not that significant. However, such customised approach is completely out of reach for the small hotels that constitute the vast majority of Booking.com business users and offer most of the rooms available to reserve.

Generally, while a number of interviewed hotels reported that they appreciate the analytics and insights that Booking.com generates for them, they lack access to some important customer information. This concerns not only contact details, but also a more detailed profiling information. In this regard, Booking.com is contrasted to Airbnb, which provides listed properties with more detailed customer profiles and additional information so that, according to an interviewee from a hotel association, “hosts are aware in advance and can serve customer better”.

The broad categories of data and analytics accessible for properties using Booking.com are summarised in the table below.

<table>
<thead>
<tr>
<th>BOOKING.COM</th>
<th>USER IDENTIFICATION DETAILS</th>
<th>DATA ON TRANSACTIONS BETWEEN BUSINESSES AND CUSTOMERS</th>
<th>BUSINESS PERFORMANCE</th>
<th>USER BEHAVIOUR</th>
<th>ANALYSES OF MARKET TRENDS/DEVELOPMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw data</td>
<td>Some (no emails!)</td>
<td>Own</td>
<td>Yes</td>
<td>No</td>
<td>Own</td>
</tr>
<tr>
<td>Analytics</td>
<td>NA</td>
<td>Competitor</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Insights/Actionable guidelines</td>
<td>NA</td>
<td>Own</td>
<td>Yes</td>
<td>Some</td>
<td>Some</td>
</tr>
</tbody>
</table>

Data provided to other entities/ open for collection

The platform claims not to share their data for the purposes of re-use with any third parties besides their business users.

Targeted internet searches did not allow to identify specific third-party data, insight of optimization tools providers building on Booking.com data and serving its business users specifically. This may be related

\textsuperscript{325} An interesting exception is the TripAdvisor instant booking product, which does provide the properties with all customer contact data. However, the company considers it unsuccessful as they did not manage to attract significant numbers of hotels.

to the fact that the platform itself provided a wide variety of analytics, insights and data tools, helping hosts optimize their listings. Scraping of Booking.com websites, however, seems to be very widespread. One indication of this is the variety of available scraping service providers, tools, tips and instructions for this on the Internet. The entities using these tools may be hotels and data product developers alike.

It is also important to mention that Booking.com offers white label solutions to other platforms. White-label means that a business can use a third-party’s software all while keeping their own branding on that page. Through these products, Booking.com provides data on its hotel inventory, to make it available to the end-users of the purchasing company. When a booking is made through such white-label software, Booking.com receives the guest personal data. However, according to the agreements between companies, any personal data related to a visitor or guest collected on the partner’s sites shall not be used by Booking.com for direct marketing purposes, unless that visitor or guest is already an existing customer of Booking.com.

**Airbnb**

Founded in 2008, Airbnb is a privately owned company is based in San Francisco, California. It provides a platform for hosts to accommodate guests with short-term lodging and tourism-related activities.

Airbnb is said to be the world’s largest accommodation provider in the sharing economy. It differs from other OTAs analysed in this paper in the sense that the majority of its business users are short-term rental providers and often private individuals, rather than hotels - although the latter (as well as the so-called mega-hosts with hundreds of listings) too widely use Airbnb. Because of the rise of the number of properties rented out for short-term accommodation in the past years, Airbnb is seen by many local governments and NGO groups as the force behind price increases in the residential housing markets, as well as many other negative externalities.

Just as Booking.com and other OTAs, Airbnb collects a lot of provided and observed data on both consumers and properties listed on the platform. Besides the data on the activities taking place on the platform, Airbnb uses third party sources to get data on references, user background information and demographics. The platform reportedly uses it in various ways - such as A/B testing, natural language processing and photo analysis - to improve its services and marketing, as well as to inform its business users.

Airbnb’s privacy policy provides a long list of third parties with which they may share the collected data. These include social media and advertising entities, providers of services related to the Airbnb Platform and the Payment Services, and Airbnb corporate affiliates. In the section on data sharing and disclosure, Airbnb claims that they share “aggregated information” about users for “industry and market analysis, research, demographic profiling, marketing and advertising, and other business purposes”, aimed at improving Airbnb’s own products and operations. In addition to this, the platform may share user data

---

330 Inside Airbnb. Available at http://insideairbnb.com/about.html
334 https://www.airbnb.com/terms/privacy_policy
with tax authorities or other public bodies\textsuperscript{335}. Meanwhile, the hosts of listings on Airbnb constitute the main category of third parties that receive platform data which they can reuse for own purposes. In the further sections, we overview them and other entities that acquire Airbnb data for re-use.

**Data provided to Airbnb business users**

As Airbnb was first launched as a sharing economy platform, establishing trust between hosts and guests has been key to its success\textsuperscript{336}. As a result, from its inception, the platform has been providing quite a lot of information for guests and hosts about each other upon booking. When a guest submits a booking request, the host can view their profile, full name, the full names of any additional guests, cancellation history, reviews of previous hosts and other information that guests agree to share. When a booking is confirmed, they also disclose additional information to assist with coordinating the trip, like phone number\textsuperscript{337}. This can be contrasted to practices of other OTAs, whose policies of not sharing guest data is subject to business user criticism.

In April 2019, Airbnb also introduced a performance dashboard for the hosts active on this platform, which is intended to help the hosts make strategic decisions. Previously, hosts were not able to easily track the performance of multiple listings, which made it challenging for them to recognize what was and was not working. More specifically, they had access to the overall outlook for the month, how many guests they have hosted and where they are originally from. The metrics also included some seasonal travel trends and information on how many people were looking at their listings\textsuperscript{338}. These were not considered detailed enough to enable hosts to better manage their business. Meanwhile, the new performance dashboard provides on-demand access to performance data across multiple listings, as well as market averages. The metrics that hosts can track using this new tool include a number of indicators that the platform deemed to be useful for the hosts (see the box below).

**BOX 21. DATA AND INSIGHTS FOR AIRBNB BUSINESS USERS**

<table>
<thead>
<tr>
<th>PERFORMANCE DASHBOARD\textsuperscript{339}</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Occupancy &amp; Rates:</strong> shows how a host’s your bookings are performing. Includes:</td>
</tr>
<tr>
<td>• Data about how the space is booked over time</td>
</tr>
<tr>
<td>• The host’s average nightly rate</td>
</tr>
<tr>
<td>• Comparisons of occupancy, average nightly rate, and revenue to a host’s market’s average</td>
</tr>
<tr>
<td>• Insights to maximize earnings for a host’s listings</td>
</tr>
<tr>
<td><strong>2. Visibility:</strong> details about the number of guests looking at a host’s listings. Includes:</td>
</tr>
<tr>
<td>• The number of times a listing is shown in search results (first page impressions) with comparison to market averages</td>
</tr>
<tr>
<td>• Listings’ views over a specific time period, with comparison to market averages</td>
</tr>
<tr>
<td>• The number of guests who click on a listing that appears in the first page of a search</td>
</tr>
</tbody>
</table>

\textsuperscript{335} An example - A pass-through registration data sharing agreement with the city of Portland is available here: http://opb-imgserve-production.s3-west-2.amazonaws.com/original/airbnb-pass-through_registration_agreement_final_and_signed_1567631972272.pdf. Information about agreements with other US cities is also available online, for instance: https://engt.co/2XXAyRu and https://www.wbur.org/bostonomix/2019/08/29/airbnb-boston-settlement-regulations


\textsuperscript{337} Airbnb Privacy Policy (2019). Available at https://www.airbnb.com/terms/privacy_policy


3. **Quality**: information about a host's guests' ratings and reviews of their listings. A host can:
   - Filter by region, listing attributes (e.g., number of bedrooms), overall or for individual listings
   - Review ratings across different time frames to identify trends
   - Compare their listings to similar listings in the area
   - Sort listings by those with the highest and lowest ratings
   - Review specific categories of guest reviews, individual listings, and guest reviews

4. **Earnings**: financial status of a host. Includes:
   - Ability to filter details by all time or for the past 12 months
   - Earnings paid out and expected earnings
   - Total nights booked
   - Total unbooked nights
   - Occupancy rate
   - Cleaning fees earned
   - Cancellation fees incurred

5. **Opportunities**: ways to optimise listings for Airbnb. Includes:
   - Superhost status or progress
   - Info around basic requirements
   - Experienced host status or progress

The platform also displays graphs showing how the host’s listings are performing compared to a competitive set of listings (e.g., how 5-star ratings for their listings compare to listings in the area). Hosts can filter by region, if their listings are in multiple regions, to increase the accuracy of the comparison. The benchmark listings for comparison are selected by Airbnb rather than the host.\(^\text{340}\)

The platform monitors how its hosts (i.e. business users) use the insights and analytics provided to measure their conversion and impacts\textsuperscript{341}. This is then used to review and improve. Nonetheless, several members of the Airbnb host community have expressed that as of mid-2019, the host analytics tools provided by the platform were not very useful. Hosts tend to complain that the platform does not consult with them on what analytics features would be useful, and does not respond when hosts reach out to it with specific requests or suggestions. Some of them mentioned specifically that they need more analytical tools to:

- See more historical data, going beyond the 12 past months.
- See real time data. If a host has an empty slot for next week, they would like to see how many people are viewing it and when they are viewing it. This would allow to adjust the pricing in an informed way, rather than guessing\textsuperscript{342}.
- Improved conversion funnel statistics, which now, according to some users, are sometimes “a bit off”\textsuperscript{343}.
- More detailed KPIs, such as “average lead time before booking”, “last three listings that have been booked in your area”, and so on.

Another specific insight that Airbnb provides is their “price tips”: suggestions to adjust hosts prices to get more bookings, based on the customer behavior patterns. Hosts see these tips as a good indication of what pricing Airbnb would expect to place the listing on the top of the search rankings\textsuperscript{344}. Nonetheless, this has received quite negative reviews from the community of Airbnb business users. They are concerned that the suggested prices are unreasonably low. For example, many hosts say that such prices would not even cover their expenses related to each stay (like cleaning, laundry, etc.)\textsuperscript{345}. Some hosts are successfully receiving bookings with much higher prices than Airbnb suggests\textsuperscript{346}. Others say that price suggestions do not correspond to the segment of their property and their targeted client (e.g., “premium” listings are receiving tips relevant to “economy range” properties). The hosts also consider the reasons for this: while some doubt the quality of the platform’s analytics algorithms, others see this as Airbnb’s strategy to attract travelers away from traditional hotels and other booking platforms with the lowest prices, “while hosts race each other to the bottom”\textsuperscript{347}. Moreover, if a listing does not book for a night, Airbnb does not make any money on it. The platform, therefore, has an incentive to suggest lower prices, because it would rather have the host book the listing for a low price than not at all\textsuperscript{348}. The lack of trust on price optimization provided by Airbnb might at least partly explain the popularity of third-party tools such as AirDNA.

Similar to other platforms in the sector, Airbnb also provides APIs for its business users, allowing them to integrate their Airbnb account with selected travel management companies or travel partners\textsuperscript{349}. The travel partners of business users can then access their booking information, including accommodation and price details. The travel partner’s responsibility for use of this information is governed by whatever


\textsuperscript{342} Airbnb Community (2019). Available at https://community.withairbnb.com/t5/Hosting/Better-analytics-for-hosts/td-p/1045502

\textsuperscript{343} “Are there any tools that people who rent on Airbnb or run a business through Airbnb wish they had, that the company isn’t providing, that would make things more efficient for them?”, Quora. Available at https://www.quora.com/Are-there-any-tools-that-people-who-rent-on-Airbnb-or-run-a-business-through-Airbnb-wish-they-had-that-the-company-isn-t-providing-that-would-make-things-more-efficient-for-them

\textsuperscript{344} Airhost Academy (n.d.). Available at https://airhostacademy.com/airbnb-price-tips/

\textsuperscript{345} Airbnb Community (2016). Available at https://community.withairbnb.com/t5/Hosting/ridiculous-price-tips/t/48793

\textsuperscript{346} “LOL! Airbnb price tips!”, Reddit. Available at https://www.reddit.com/r/AirBnB/comments/9hcc81/lol_airbnb_price_tips/

\textsuperscript{347} Airbnb Community (2016). Available at https://community.withairbnb.com/t5/Hosting/ridiculous-price-tips/t/48793

\textsuperscript{348} Airhost Academy (n.d.). Available at https://airhostacademy.com/airbnb-price-tips/


111
agreements or contracts business users have with them. The business user is the data controller for the trip data of its travelers, and is responsible for the partner’s handling of the data.\textsuperscript{350}

**TABLE 15. TYPES OF DATA AND ANALYTICS ACCESSIBLE BY PROPERTIES LISTED ON AIRBNB**

<table>
<thead>
<tr>
<th>AIRBNB USER IDENTIFICATION DETAILS</th>
<th>DATA ON TRANSACTIONS BETWEEN BUSINESSES AND CUSTOMERS</th>
<th>BUSINESS PERFORMANCE</th>
<th>USER BEHAVIOUR</th>
<th>ANALYSES OF MARKET TRENDS/DEVELOPMENT S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Own</td>
<td>Competitor</td>
<td>Own</td>
<td>Competitor</td>
<td>Own</td>
</tr>
<tr>
<td>Raw data</td>
<td>Yes (profile data)</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Analytics</td>
<td>NA</td>
<td>Some</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Insights/Actionable guidelines</td>
<td>NA</td>
<td>Some</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Data provided to other entities/ open for collection

As in the case of online platforms in general, the nature of Airbnb business requires that a lot of its data are made publicly accessible. In order to meet its goal of connecting hosts with guests, the Airbnb platform must display information about the nature and size of available units, their price, general location, and general availability. The site also contains photographs of units, their verbal descriptions, amenities and location, information about hosts and reviews (which may contain information about guests and what they did during their stay, information about host and guest interactions; information about the unit, about the dates of the stay, and about the area in which the unit is located). Although the primary purpose of these data is to provide information about specific units that are for rent, taken together and subject to analytics, they can provide rich information about the number of rental units available in particular areas, the availability of those units over time, their price (varying by time of year and over time), and more.\textsuperscript{351}

These data are used in analytics to assist those seeking to enter the market and to appropriately price their rental units. Companies such as AirDNA and others, serving the current and potential OTA business users with data and analytics, have built their products and services on this data, which is constantly scraped from the platform’s websites and collected from hosts who opted in. Airbnb hosts that use AirDNA services claim that the analytics this company generates on their properties are reasonably accurate.\textsuperscript{352} This illustrates the wide possibilities to use publicly available data to extrapolate meaningful and precise insights.

What concerns the interest in Airbnb data by public authorities, civil society and journalists, the platform provides and especially interesting case because of its business model. Given that a large share of Airbnb hosts provide accommodation in residential properties rather than hotels, it raises questions about, among other things, the platform’s effects on the cost and availability of long term accommodation, its impact on incumbent short-term accommodation providers, the incidence of discrimination in Airbnb rentals and pricing, and the extent to which the platform is used to support full scale commercial ventures. In various


\textsuperscript{352} “Is the data provided by airdna accurate?”, Quora. Available at https://www.quora.com/Is-the-data-provided-by-airdna-accurate
research publications, authors argue that it is in public interest for Airbnb data (along with data from other “sharing economy” platforms) to be accessible for broader audiences. Currently only Airbnb’s publicly available information can be used to understand the extent to which units are really just “excess space” or are diverted from the pool of available long-term accommodation. Data scraping, therefore, remains a primary source of such data for these groups of users in the absence of adequate voluntary data-sharing by the company. However, Airbnb has joined the initiative to share short-term rental data with Eurostat. This should enable governments, tourism boards, journalists and NGOs to get a more convenient access to this data.

### Expedia

First launched in 1996 in the US, Expedia is an online travel agency and metasearch engine. It is currently owned and operated by Expedia Group, ranked first on the global list of top earning travel companies. Other brands of the group include Hotwire, CarRentals.com, Hotels.com, HomeAway and Trivago, among others.

As an OTA, Expedia also heavily emphasizes the importance of data for its business model and activities. It uses third-party travel industry research, combined with first-party travel data collected across the network of Expedia’s sites to continuously improve and optimize.

According the Expedia’s privacy policy, the platform shares the data they collect with:

- Business users (“Suppliers”) such as hotels, airlines, car-rental companies, insurance, vacation-rental property owners and managers, and where available, activity providers, rail, or cruise lines who fulfill the booking.
- Expedia Group companies.
- Third-party service providers supporting the delivery of services and the operation of platform’s business (for example, to provide credit card processing, customer service, business analytics, and fraud prevention and compliance services, advertising). These are not allowed to use this information for purposes of their own direct marketing.
- Business partners, such as a third-party loyalty program for which consumers could earn points by completing a booking on Expedia.
- Social media, when the consumer access certain features like Facebook’s “Like” button or a single sign-on that allows you to login with your social media credentials.
- In some cases, as required by law, or to stop illegal activities, with government agencies.

The platform shares data in order to carry out business transactions and ensure a proper functioning of its services. Expedia does not share data with third parties - except its business users - for them to re-use the data for own analytics, service improvement or product development.

### Data provided to Expedia business users

As the other OTAs overviewed above, Expedia actively invests in developing and providing insights for its business users. In 2016, Expedia launched Rev+, a revenue management tool for its business users, designed to provide them with actionable data and insights. Since then, the tool is continuously updated

---


with new functionalities. The key features of Rev+ allow hotels to gauge their rates against competitors over the course of 90 days, alongside other market factors; review forecasted demand for a market, based on data points captured across the portfolio of Expedia Inc. brands; and monitor their property’s revenue performance, guest behavior, and market outlook against the property’s Competitive Set. The metrics are presented together with specific pointers on how they could be used, and what they can help achieve, for example: more visibility on the platform, better conversion rates and targeting high-value guests. This is accompanied by customized tips and recommendations. More details on the metrics provided are listed in the box below. Rev+ comes at no additional costs for hotels, and it does not require additional sign-up.

**BOX 22. ANALYTICS PROVIDED TO EXPEDIA’S BUSINESS USERS**

**EXPEDIA’S REV+**

**Competitive set:** the business users can customize their competitive set and see how they compare in terms of different metrics. The set can include from 5 to 20 properties, including their own, that can be seen on a map. Based on this information, hotels, among other things, can evaluate their Competitive Set’s pricing strategy in different markets, and adjust their pricing strategy.

**Price Calendar:** a comprehensive view of a hotel’s pricing relative to their competitive set, including a possibility to view your property’s pricing data by any seven days, a full month or the next 365 days.

**Price Optimization:** price alerts, showing which days the price set by the hotel may be too high or too low, in comparison to its competitive set.

---


358 Ibid.
**Market Demand Score**: hotels can view future demand for their market based on millions of data points from the Expedia, Inc. family of websites and its affiliations.

**Market Occupancy Forecast**: estimates of how many rooms will be sold in a destination, on a specific date, in the future.

**Property analytics**: shows property’s visibility, conversion and guest value. Could be compared against the property’s last year’s performance or a competitive set of other business users on the platform (their averages).
Guest insights: shows guests’ countries, types of trips and who they travel with for the property and its Competitive set, allowing to better understand the booking behavior and recognise opportunities for tailored strategies.

The Rev+ analytics could be accessed through Expedia Partner Central portal or Partner Reporting Suite. The latter provides a self-service access to various datasets and pdf reports (e.g. on the property’s year-to-date room nights, average rates, comparisons with competitive sets, performance trends, monthly states, and booking trends by regions, segments and other booking details359) based on the presented metrics360. The hotels can also subscribe to a notification system, which alerts hotels to changes to rates over the course of the last 24 hours.

The platform representatives claim they are actively collecting feedback from the hotels to better understand what they like in the analytics tools, as well as where they see opportunities for

improvement\(^{361}\). The analytics seem to be appreciated in the hotelier community, as it is detailed, presented in an intuitive way, and immediately applicable in decision-making.

The main incentive for Expedia to continuously invest in host analytics and their improvement is to keep their business users satisfied in the competitive OTA market so that they do not switch, and allow them to generate more bookings, which are the source of Expedia’s revenues.

The platform, as other OTAs, do not provide any raw data, as it “wouldn’t make sense”, neither from Expedia’s business perspective, nor for the hotels, most of which are small enterprises without the necessary analytical capacities. Moreover, provision of aggregated insights rather than granular data, according to the interviewed representatives, allow the platform to strike right balance between complying with legal requirements, GDPR, antitrust and providing necessary insight to partners.

In addition to the Rev+ tools and metrics, Expedia provides it business users with a resource centre, containing tips and other materials for more effective revenue management\(^{362}\), distribution strategy, marketing and guest relations\(^{363}\). The platform also offers an API with “rich and dynamic property content” on listings on Expedia\(^{364}\).

What concerns data and information about customers, the platform provides the customer data needed to process payments, including the name and address. The hotels are left with the possibility to collect more detailed personal data at check-in. As in case of Booking.com, not being able to access more detailed customer personal and contact data is one of the key complains of the platform’s business users. However, the representatives of hotel associations noted in an interview that due to its smaller market share in Europe Expedia is more flexible in terms of customer data provision than Booking.com (and this is likely the opposite in the US in which Expedia dominates).

### TABLE 16. TYPES OF DATA AND ANALYTICS ACCESSIBLE BY PROPERTIES LISTED ON EXPEDIA

<table>
<thead>
<tr>
<th>EXPEDIA USER IDENTIFICATION DETAILS</th>
<th>DATA ON TRANSACTIONS BETWEEN BUSINESSES AND CUSTOMERS</th>
<th>BUSINESS PERFORMANCE</th>
<th>USER BEHAVIOUR</th>
<th>ANALYSES OF MARKET TRENDS/DEVELOPMENT S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Own</td>
<td>Competitor</td>
<td>Own</td>
<td>Competitor</td>
</tr>
<tr>
<td>Raw data</td>
<td></td>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Analytics</td>
<td>NA</td>
<td>Some</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Insights/Actionable guidelines</td>
<td>NA</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Data provided to other entities/ open for collection**

The platform argues that it does not share their data for the purposes of re-use with any third parties besides their business users.

---

363 https://welcome.expediagroup.com/en/resources
364 https://expediapartnersolutions.com/products/api
The platform representatives mentioned that they are aware of the scraping activities on Expedia’s sites, but they did not see this as an issue. They did not know if any optimization or insight providers servicing the business users of Expedia are engaged in such activities, or if any such providers exist at all (although AirDNA focuses on Expedia’s brand HomeAway offering short-term rentals). Targeted Google searches also did not allow to identify such third-party data providers focusing specifically on Expedia and its data.

**TripAdvisor**

Founded in 2000 (and split from Expedia in 2011), TripAdvisor is an online travel company that operates a global platform with user-generated content, price comparison tools, and online reservations for transportation, lodging, travel experiences, and restaurants. Its business model, however, differs from other major OTAs: most of the hotel-related services of TripAdvisor operate as advertising and redirecting tools rather than a marketplace. Booking.com and Expedia are among the main clients of these TripAdvisor services.

TripAdvisor’s has also introduced a hotel Instant Book tool a few years ago. This part of TripAdvisor business functions similar to the other OTAs: consumers use the platform to book rooms through TripAdvisor, rather than being directed to Expedia or Booking.com though their paid ads (as in the main part of TripAdvisor business). However, according to the interviewed representative of the platform, this business did not take off as hotels are reluctant to participate and the commission on these services generate a very low share of revenues. This differs quite drastically from the other OTAs analysed in this paper, for which the intermediation services are the key source of revenues. Direct bookings generate more sales in the platform in restaurant and attraction segments. Nonetheless, these commission fees also contribute to only around 20% of TripAdvisor revenues.

According to the platform’s privacy policy, the information it collects on its registered and unregistered users (consumers) differs. The platform’s representative explained only around 30% of the platform’s users are registered, while the rest 70% are not. This is related to the fact that TripAdvisor’s key value added for consumers is travel advice rather than booking of travel-related services, for which consumers must be registered. From unregistered users, TripAdvisor collects IP addresses, and track the behaviour on its sites using cookies with no personally identifiable information. Meanwhile, registered users provide their data while registering, leaving reviews, and engaging in discussions in forums. TripAdvisor is able to track the digital footprint on its websites more comprehensively only from the registered users. The platform mostly uses customer data for advertising purposes, as it enables to display personalised and more relevant ads, with better conversion rates for the platform’s advertising partners.

According to its privacy policy, the platform provides this information to several types of third parties:

- TripAdvisor “suppliers” - business users of the platform’s intermediation services.
- TripAdvisor business partners who advertise on the site, when the users “click-through”.
- TripAdvisor parent company, subsidiaries and sister companies.
- Third parties specifically engaged by TripAdvisor to provide services to TripAdvisor (such as market research firms, marketing services providers and advertising agencies), for the purpose of providing the specified services to TripAdvisor (rather than re-use).

---


Public authorities for law enforcement purposes\textsuperscript{367}.

Tripadvisor claims that they do not sell or transfer any personal information to any third party, other than their partners, as described above\textsuperscript{368}.

**Data provided to TripAdvisor business users**

The level of access to data and analytics for TripAdvisor’s business users depends on the types of their contract with the platform. Every property (e.g., restaurant, hotel or a tour agency) listed for free on TripAdvisor receive a number of metrics on how their business is doing on the platform, including benchmarking with two other establishments (what is the most comparable business is decided by TripAdvisor), and number of reviews. The business users are also provided with tips how to manage their reviews.

Businesses who opt in for the paid Premium service package (called “Business advantage” contract; according to platform’s representative, a minority of hotels listed are using this), among other upgrades in the service package (e.g., possibility to upload more photos), also get more options for analytics, and the ability to decide the set of establishments that they want to be compared to. Owners can build and modify up to 10 competitor sets and compare their own reviews and rankings to that of their competitors over time\textsuperscript{369}.

The data is provided through business user account management page, in graphs and dashboards (first launched in 2015\textsuperscript{370}). Data is viewable in weekly, monthly or quarterly displays. The business users can also download some of TripAdvisor data to analyse their performance off the platform, and use those insights as a marketing tool.

For hotels in the TripAdvisor’s Instant Book programme, the platform also transfers all the customer data, including name, address, payment, information, and email address once the payment is made. This differentiates TripAdvisor from OTAs, who disclose only the minimum of this information necessary to process the transaction. Despite this advantage, TripAdvisor’s Instant Book was not seen by the hotels as an alternative to traditional OTAs.

In addition to this, the platform has its TripAdvisor Insights\textsuperscript{371} portal, in which it provides various research results that can help their business users improve, but is accessible to broader audiences as well. The reports provided there are mostly based on TripAdvisor end-customer surveys and study results of other organisations (e.g., Oxford Economics Global Travel Market Study, comScore Worldwide Path to Purchase Report, etc.). In addition to that, the platform offers its business users various evidence-based tips of how to use the opportunities that TripAdvisor provides more effectively: for instance, that replies to customer reviews matter, what best ways to reply to positive and negative reviews are, what tourists from different regions expect when they use accommodation services, and so on.

Nonetheless, the interviewed representative of TripAdvisor noted that although they are willing to provide more data, analytics and insights to their business users, the level of use of these data and insights

---


\textsuperscript{368} Tripadvisor Privacy Policy (2020). Available at https://www.tripadvisor.com/pages/privacy_pre_060407.html


\textsuperscript{371} https://www.tripadvisor.com/TripAdvisorInsights
among businesses “is just too low”, although a notable difference between small and large establishments exists. Large hotel chains use the data in, for example, their marketing strategies, and have specific staff and departments engaged in these activities. Therefore, they sometimes hear requests for access to more consumer data. Meanwhile, independent hotels or restaurants make use of the analytics or recommendations provided by the platform (e.g. about replying to reviews).

**TABLE 17. TYPES OF DATA AND ANALYTICS ACCESSIBLE BY PROPERTIES LISTED ON TRIPADVISOR**

<table>
<thead>
<tr>
<th>TRIP ADVISOR</th>
<th>USER IDENTIFICATION DETAILS</th>
<th>DATA ON TRANSACTIONS BETWEEN BUSINESSES AND CUSTOMERS</th>
<th>BUSINESS PERFORMANCE</th>
<th>USER BEHAVIOUR</th>
<th>ANALYSES OF MARKET TRENDS/DEVELOPMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw data</td>
<td>Own</td>
<td>Own</td>
<td>Own</td>
<td>Own</td>
<td>NA</td>
</tr>
<tr>
<td>Analytics</td>
<td>NA</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>NA</td>
</tr>
<tr>
<td>Insights/Actionable guidelines</td>
<td>NA</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Data provided to other entities/ open for collection**

Similar to OTAs, the data provided to businesses other than the platform’s business users is limited. The platform claims that they do not provide data for Skift and other similar travel insight companies for reuse. As a public company (listed in New York Stock Exchange), TripAdvisor is also obliged to provide its main financial data publicly. Travel data companies Skift and Phocuswright, according to the interviewed Tripadvisor representative, use this type of data to prepare their insights about the travel sector.

The platform does not see the scraping of its websites as an issue. TripAdvisor’s representative did not know any insight of optimization companies that provide data services to the platform’s business users. Targeted searches on Google also did not allow to identify any optimization or insight companies providing their services specifically to Tripadvisor’s business users.

**eDreams ODIGEO**

Formed in 2011, eDreams ODIGEO is one of the largest European OTA businesses based in Barcelona, Spain. The company operates five brands: Opodo, eDreams, Go Voyages, Travellink and Liligo – offering services of travel booking, from flights to accommodation and car rental. eDreams ODIGEO is listed on the Spanish Stock Market.372 With a presence in 46 markets, it is Europe’s largest flight retailer, serving over 18 million customers as of 2019. It therefore differs from other OTAs reviewed in this paper as their main business users are airlines, and over 80% of their revenues are generated through flight sales. However, eDreams ODIGEO offers intermediation services to hotels and short-term rental providers as well[373], using a white label technology solution that it purchases. This enables eDreams to offer package holidays (flight and accommodation) to consumers.

372 https://www.edreamsodigeo.com/about-us/
373 https://www.odigeoconnect.com/en/
Besides active (e.g. through the creation of an account, consumer surveys, communication with the platform) and passive/automatic data collection on consumers and business users, eDreams ODIGEO discloses that they periodically obtain personal information about customers from business partners and other independent third-party sources (e.g. contact information such as email, purchases or demographic information).

The platform uses anonymised or pseudonymised personal consumer data for analytical purposes, to enhance user experience, for testing and troubleshooting to improve the functionality and quality of online travel services, as well as to detect of fraud and security.

In its privacy policy, the eDreams ODIGEO states that it shares the data collected on customers with several types of third parties:

- Travel partners (i.e., the business users), including airlines, hotels, car rental companies, and so on.
- eDreams ODIGEO group companies.
- Third party service providers (e.g. those providing IT and hosting services, customer support, analytics, payment and financial service providers for chargeback, fraud detection, prevention purposes, etc.).
- Business partners that provide some of the website or app services could fully or partially.
- Social media providers, when consumers log in with social media, click on a social media “like” button integrated in the website or app by plugins, or use any social media services to interact with the platform.
- Competent authorities for the purposes of law enforcement.

Based on the available evidence, only business users are provided with data for the purposes of their own business improvement (i.e. re-use). In other cases, data is transferred to other parties mostly to support eDreams own business processes.

**Data provided to eDreams business users**

Given the specificities of the aviation industry related to the highest security requirements, consumers booking a flight must provide detailed personal information, including their passport or ID number. When the booking is done via eDreams, this information is automatically transferred to the airlines. However, what concerns data on consumer behaviour on the platform (e.g., how they search for flights, what offset they tend to choose, etc.), its provision to the airlines is limited.

For the hotels using eDreams OTA services, the platform provides an Analytics section in the eDreams ODIGEO Connect Extranet. It is aimed at providing insights and data to take better decisions and improve hotel revenue. The section provides various information organized into easy-to-read charts including, for example, business users’ bookings and room nights per month (by stay or by check-in date), as well as insights on traveler groups and destinations. Some of the insights that the platform provides is aimed at helping hoteliers to forecast their demand. Moreover, the platform provides hotels with insights about the platform’s flight customers, which should help the properties to target this type of travellers.

---

374 eDream Privacy Policy. Available at https://www.edreams.com/privacy-policy/
375 eDream Privacy Policy. Available at https://www.edreams.com/privacy-policy/
376 eDream Privacy Policy. Available at https://www.edreams.com/privacy-policy/
In addition to this, the platform periodically publishes traveler insight reports and tips for their business users (e.g., how to manage consumer reviews, take better photos or improve hotel websites). These services are somewhat similar to those offered by TripAdvisor. These are publicly available on one of its websites\(^\text{379}\).

**TABLE 18. TYPES OF DATA AND ANALYTICS ACCESSIBLE BY EDREAMS BUSINESS USERS**

<table>
<thead>
<tr>
<th>EDREAMS USER IDENTIFICATION DETAILS</th>
<th>DATA ON TRANSACTIONS BETWEEN BUSINESSES AND CUSTOMERS</th>
<th>BUSINESS PERFORMANCE</th>
<th>USER BEHAVIOUR</th>
<th>ANALYSES OF MARKET TRENDS/ DEVELOPMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw data</td>
<td>Own</td>
<td>Yes</td>
<td>Own</td>
<td>Yes</td>
</tr>
<tr>
<td>Analytics</td>
<td>NA</td>
<td>Some</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Insights/Actionable guidelines</td>
<td>NA</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Data provided to other entities/open for collection

Interviewed representative of eDreams distinguished between two types of third-party companies with whom the platform shares data:

— White label technology solutions that eDreams purchase (e.g., for hotel bookings or meta-search facilities). Using these solutions implies continuous data exchanges between eDreams and the partners. Once a booking is made on a white label solution, its operator receives personal customer data to complete the transaction. However, based on bilateral agreements, there are limitations on how this data may be used. The partners cannot reuse this data for own purposes such as commercial promotional personalization or marketing.

— Providers of marketing and advertising services. As other platforms in the tourism sector, eDreams uses Google and Facebook advertising services extensively, which means it must provide them with significant amounts of data. Given their market position, the advertising companies enforce their standard terms and conditions, which do not prohibit them to use eDreams data for whatever purposes they wish to. This is seen problematic not only by eDreams, but also by many OTAs, as Facebook and Google can use these data to develop their own verticals competing in the online travel sector.

Specific entities that scrape eDreams websites, or use data collected this way, were not identified (interestingly, eDreams itself is known for scraping airline websites\(^\text{380}\)). Targeted Internet searches did not allow to identify any data or optimization companies focusing on eDreams data or information on the specific uses of data scraped from this OTA.


References

21. eBay User Privacy Notice: https://www.ebay.com/help/policies/member-behaviour-policies/user-privacy-notice-privacy-policy?id=4260#section4
45. Data Streams, (2018), “The prisoner’s dilemma of data sharing (and how to solve it)”, *Data Streams*. Available at https://www.datastreams.io/dilemma-data-sharing-platform-datastreams/
46. Dayton, E. (2020), „Amazon statistics you should know: opportunities to make the most of America’s top online marketplace”, *Big Commerce*. Available at https://www.bigcommerce.com/blog/amazon-statistics/a-shopping-experience-beyond-compare
64. eDreams Privacy Policy. Available at https://www.edreams.com/privacy-policy/
Analytical paper #5


# List of interviewees

<table>
<thead>
<tr>
<th>#</th>
<th>INTERVIEWEE/ ORGANISATION</th>
<th>INTERVIEWEE TYPE</th>
<th>RESPONSE/ STATUS</th>
<th>INTERVIEW DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEW INTERVIEWS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>E-COMMERCE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Amazon</td>
<td>Product marketplace</td>
<td>Completed</td>
<td>27/04/2020</td>
<td></td>
</tr>
<tr>
<td>2. Allegro</td>
<td>Product marketplace</td>
<td>Completed</td>
<td>15/04/2020</td>
<td></td>
</tr>
<tr>
<td>3. eBay</td>
<td>Product marketplace</td>
<td>Completed</td>
<td>10/04/2020</td>
<td></td>
</tr>
<tr>
<td>4. Rakuten</td>
<td>Product marketplace</td>
<td>Completed</td>
<td>26/03/2020</td>
<td></td>
</tr>
<tr>
<td>5. Real.de</td>
<td>Product marketplace</td>
<td>Completed</td>
<td>17/03/2020</td>
<td></td>
</tr>
<tr>
<td>6. Zalando</td>
<td>Product marketplace</td>
<td>Completed</td>
<td>18/03/2020</td>
<td></td>
</tr>
<tr>
<td>7. E-commerce seller 1</td>
<td>Business user (Amazon)</td>
<td>Completed</td>
<td>06/03/2020</td>
<td></td>
</tr>
<tr>
<td><strong>ONLINE TRAVEL AGENCIES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. EU Travel Tech</td>
<td>Industry association (platforms)</td>
<td>Completed</td>
<td>13/03/2020</td>
<td></td>
</tr>
<tr>
<td>9. European Hotel Forum</td>
<td>Industry association</td>
<td>Completed</td>
<td>29/04/2020</td>
<td></td>
</tr>
<tr>
<td>10. Hospitality Europe</td>
<td>Industry association</td>
<td>Completed</td>
<td>17/04/2020</td>
<td></td>
</tr>
<tr>
<td>11. Expedia Group</td>
<td>OTA</td>
<td>Completed</td>
<td>10/03/2020</td>
<td></td>
</tr>
<tr>
<td>12. TripAdvisor</td>
<td>OTA</td>
<td>Completed</td>
<td>27/03/2020</td>
<td></td>
</tr>
<tr>
<td>13. eDreams ODIGEO</td>
<td>OTA</td>
<td>Completed</td>
<td>06/05/2020</td>
<td></td>
</tr>
<tr>
<td><strong>APP STORES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. APP Association (ACT)</td>
<td>Industry association (SMEs)</td>
<td>Completed</td>
<td>07/04/2020</td>
<td></td>
</tr>
<tr>
<td>15. Developers Alliance</td>
<td>Industry association</td>
<td>Completed</td>
<td>18/03/2020</td>
<td></td>
</tr>
<tr>
<td>16. Computer &amp; Communications Industry Association</td>
<td>Industry association</td>
<td>Completed</td>
<td>19/03/2020</td>
<td></td>
</tr>
<tr>
<td>17. Google Play</td>
<td>App Store</td>
<td>Completed</td>
<td>01/04/2020</td>
<td></td>
</tr>
<tr>
<td>18. Apple App Store</td>
<td>App Store</td>
<td>Completed</td>
<td>03/04/2020</td>
<td></td>
</tr>
<tr>
<td><strong>EXPERTS AND ORGANISATIONS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. Silvia Martinelli</td>
<td>Expert</td>
<td>Completed</td>
<td>24/03/2020</td>
<td></td>
</tr>
<tr>
<td>20. Teresa Scassa</td>
<td>Expert</td>
<td>Completed</td>
<td>16/03/2020</td>
<td></td>
</tr>
<tr>
<td>21. Heiko Richter and Peter Slowins</td>
<td>Expert</td>
<td>Completed</td>
<td>20/04/2020</td>
<td></td>
</tr>
<tr>
<td>22. Centre for Data Innovation</td>
<td>Think tank</td>
<td>Completed</td>
<td>19/03/2020</td>
<td></td>
</tr>
<tr>
<td>23. European Tech Alliance (EUTA)</td>
<td>Industry alliance</td>
<td>Completed</td>
<td>15/04/2020</td>
<td></td>
</tr>
<tr>
<td>24. European Digital SME Alliance</td>
<td>Association</td>
<td>Completed</td>
<td>10/04/2020</td>
<td></td>
</tr>
</tbody>
</table>

**INTREVIEWS COMPLETED FOR OTHER ANALYTICAL PAPERS**

<table>
<thead>
<tr>
<th>#</th>
<th>INTERVIEWEE/ ORGANISATION</th>
<th>INTERVIEWEE TYPE</th>
<th>RESPONSE/ STATUS</th>
<th>INTERVIEW DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-COMMERCE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25. Amazon</td>
<td>Product marketplace</td>
<td>Completed</td>
<td>03/12/2019</td>
<td></td>
</tr>
</tbody>
</table>

133
<table>
<thead>
<tr>
<th>#</th>
<th>INTERVIEWEE/ ORGANISATION</th>
<th>INTERVIEWEE TYPE</th>
<th>RESPONSE/ STATUS</th>
<th>INTERVIEW DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>26.</td>
<td>Allegro</td>
<td>Product marketplace</td>
<td>Completed</td>
<td>06/12/2019</td>
</tr>
<tr>
<td>27.</td>
<td>eBay</td>
<td>Product marketplace</td>
<td>Completed</td>
<td>17/10/2019</td>
</tr>
<tr>
<td>29.</td>
<td>Real.de</td>
<td>Product marketplace</td>
<td>Completed</td>
<td>16/11/2019</td>
</tr>
<tr>
<td>30.</td>
<td>Zalando marketplace</td>
<td>Product marketplace</td>
<td>Completed</td>
<td>20/11/2019</td>
</tr>
<tr>
<td>31.</td>
<td>E-commerce seller 1</td>
<td>Business user</td>
<td>Completed</td>
<td>07/11/2019</td>
</tr>
<tr>
<td>32.</td>
<td>E-commerce seller 2</td>
<td>Business user</td>
<td>Completed</td>
<td>18/10/2019</td>
</tr>
<tr>
<td>33.</td>
<td>E-commerce seller 3</td>
<td>Business user</td>
<td>Completed</td>
<td>20/11/2019</td>
</tr>
<tr>
<td>34.</td>
<td>E-commerce seller 4</td>
<td>Business user</td>
<td>Completed</td>
<td>26/11/2019</td>
</tr>
<tr>
<td>35.</td>
<td>E-commerce seller 5</td>
<td>Business user</td>
<td>Completed</td>
<td>11/11/2019</td>
</tr>
<tr>
<td>36.</td>
<td>E-commerce seller 6</td>
<td>Business user</td>
<td>Completed</td>
<td>25/11/2019</td>
</tr>
<tr>
<td>37.</td>
<td>E-commerce seller 7</td>
<td>Business user</td>
<td>Completed</td>
<td>04/12/2019</td>
</tr>
<tr>
<td>38.</td>
<td>E-commerce seller 8</td>
<td>Business user</td>
<td>Completed</td>
<td>02/12/2019</td>
</tr>
<tr>
<td>39.</td>
<td>E-commerce seller 9</td>
<td>Business user</td>
<td>Completed</td>
<td>21/11/2019</td>
</tr>
<tr>
<td>40.</td>
<td>E-commerce seller 10</td>
<td>Business user</td>
<td>Completed</td>
<td>18/12/2019</td>
</tr>
<tr>
<td>41.</td>
<td>Algopix</td>
<td>Optimisation company</td>
<td></td>
<td>23/10/2019</td>
</tr>
<tr>
<td>42.</td>
<td>Jungle Scout</td>
<td>Optimisation company</td>
<td></td>
<td>08/11/2019</td>
</tr>
<tr>
<td></td>
<td><strong>ONLINE TRAVEL AGENCIES</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43.</td>
<td>Booking.com</td>
<td>OTA</td>
<td>Completed</td>
<td>08/11/2019</td>
</tr>
<tr>
<td>44.</td>
<td>Expedia Group</td>
<td>OTA</td>
<td>Completed</td>
<td>19/11/2019</td>
</tr>
<tr>
<td>45.</td>
<td>TripAdvisor</td>
<td>OTA</td>
<td>Completed</td>
<td>13/11/2019</td>
</tr>
<tr>
<td>46.</td>
<td>AirDNA</td>
<td>Optimisation company</td>
<td>Completed</td>
<td>07/11/2019</td>
</tr>
<tr>
<td>47.</td>
<td>Accommodation service provider 1</td>
<td>Small hotel</td>
<td>Completed</td>
<td>13/11/2019</td>
</tr>
<tr>
<td>48.</td>
<td>Accommodation service provider 2</td>
<td>Small hotel</td>
<td>Completed</td>
<td>27/11/2019</td>
</tr>
<tr>
<td>49.</td>
<td>Accommodation service provider 3</td>
<td>Small hotel</td>
<td>Completed</td>
<td>18/11/2019</td>
</tr>
<tr>
<td>50.</td>
<td>Accommodation service provider 4</td>
<td>Large hotel</td>
<td>Completed</td>
<td>09/12/2019</td>
</tr>
<tr>
<td>51.</td>
<td>Accommodation service provider 5</td>
<td>Large hotel</td>
<td>Completed</td>
<td>06/12/2019</td>
</tr>
<tr>
<td>52.</td>
<td>Accommodation service provider 6</td>
<td>Large hotel</td>
<td>Completed</td>
<td>04/12/2019</td>
</tr>
<tr>
<td>53.</td>
<td>Accommodation service provider 7</td>
<td>Large hotel</td>
<td>Completed</td>
<td>09/12/2019</td>
</tr>
<tr>
<td></td>
<td><strong>APP STORES</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54.</td>
<td>Apple (App store, iOS, sdk)</td>
<td>App store</td>
<td>Completed</td>
<td>30/10/2019</td>
</tr>
<tr>
<td>55.</td>
<td>Google (Play store, Android, Fuchsia)</td>
<td>App store</td>
<td>Completed</td>
<td>31/10/2019; 05/12/2019</td>
</tr>
<tr>
<td>56.</td>
<td>Apptopia</td>
<td>Optimisation company</td>
<td>Completed</td>
<td>24/10/2019</td>
</tr>
<tr>
<td>57.</td>
<td>App developer 1</td>
<td>Large app developer</td>
<td>Completed</td>
<td>07/10/2019</td>
</tr>
<tr>
<td>58.</td>
<td>App developer 2</td>
<td>Small app developer</td>
<td>Completed</td>
<td>25/09/2019</td>
</tr>
<tr>
<td>59.</td>
<td>App developer 3</td>
<td>Large app developer</td>
<td>Completed</td>
<td>11/10/2019</td>
</tr>
<tr>
<td>60.</td>
<td>App developer 4</td>
<td>Medium app developer</td>
<td>Completed</td>
<td>11/09/2019</td>
</tr>
<tr>
<td>61.</td>
<td>App developer 4</td>
<td>Medium app developer</td>
<td>Completed</td>
<td>17/09/2019</td>
</tr>
</tbody>
</table>
Interview topics

1. What data, collected and held by platforms, is important for their business users and other businesses active in specific sectors?
   - What are the (potential) users of platform data besides the businesses operating on them and their customers, e.g. other platforms?
   - What purposes do they need platform data for?
   - What types of data do these players need for these purposes and in which form (raw data, data analytics, etc.)?

2. What kinds of data do platforms provide/open for each type of identified users? How and under what arrangements?
   - Is there evidence on discriminatory access to data and exclusivity arrangements?
   - Have practices evolved over time, e.g. has a previously more open access been restricted and if yes for which reasons?

3. What types of data, important to different users, do platforms refuse to share and on what grounds?

4. What are the incentives and constraints for platforms to share data?
   - What are the economic incentives and constraints?
   - What are the legal incentives and constraints?
   - What are the technical incentives and constraints (e.g. portability and interoperability)?
   - How could data sharing be facilitated with a view to achieve more competition and innovation?
   - Are there other incentives and constraints?

5. Is the needed data which is not provided by platforms:
   - Indispensable for specific purposes?
   - Replicable?
   - Do alternative sources of such data exist?
   - If yes, under what conditions?

6. What are the possible solutions to address platform refusals to share data important to other users?
   - What are the existing practices (policy, self-regulation, role of data brokers etc.)?
   - What are possible options?
   - Could access to data under fair, reasonable and non-discriminatory (FRAND) conditions be mandated?
   - What are the risks and opportunities related to each option?
   - What would be the case for a “reversal of the burden of proof” for data sharing and which benefits could be derived from it?