Platform data access and secondary data sources

Analytical paper 1
Authors:

Vaida Gineikytė (PPMI), Dr Egidijus Barcevičius (PPMI), Loreta Matulevič (PPMI)

Contact:

Dr Egidijus Barcevičius, project manager
PPMI
Gedimino pr. 50
Vilnius, 01110 Lithuania
egidijus.barcevicius@ppmi.lt
Tel: + 370 5 249 7056


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 ........................................................................................................... 3
Setting the context ............................................................................................................. 3
Sources and methodology ................................................................................................. 4
Business users: data needs and access to data ................................................................. 4
The role of data companies ............................................................................................... 6
Obliging platforms to share data ........................................................................................ 8
Further research ................................................................................................................. 8

1. Introduction .................................................................................................................. 10

2. Scope and methodology ............................................................................................... 13
   2.1. Scope of the analytical paper ................................................................................ 13
   2.2. Methodology of data collection and analysis ......................................................... 14

3. Platform business user data needs .............................................................................. 17
   3.1. Types of data for platform business users .............................................................. 17
   3.2. Sector-specific data needs ...................................................................................... 21
      3.2.1. E-commerce sellers ......................................................................................... 22
      3.2.2. App developers .............................................................................................. 24
      3.2.3. Accommodation and hospitality industry ...................................................... 25
   3.3. General data needs of platform business users ...................................................... 26

4. Platform data and business user access ..................................................................... 27
   4.1. Amazon Marketplace ............................................................................................ 28
   4.2. eBay ...................................................................................................................... 34
   4.3. Google Play app store ........................................................................................... 36
   4.4. Booking.com ......................................................................................................... 39
   4.5. Reasons for providing data or restricting the access .............................................. 43

5. Business user perspective on data access .................................................................... 46
   5.1. Business users’ perspective on platform data access ........................................... 46
   5.2. Data alternatives for business users ..................................................................... 49

6. Data companies .......................................................................................................... 53
   6.1. Types of data supplied by data brokers .................................................................. 58
      6.1.1. E-commerce .................................................................................................... 62
      6.1.2. Software development .................................................................................... 63
      6.1.3. Accommodation and hospitality ................................................................... 64
      6.1.4. Non sector-specific data providers ................................................................. 64
   6.2. Data sources and collection methods ..................................................................... 65
   6.3. Data management and protection ......................................................................... 69

7. Conclusions and discussion ......................................................................................... 73

List of interviewees ......................................................................................................... 76
Reference list .................................................................................................................... 78
LIST OF FIGURES

Figure 1. Total completed interviews according to their type on the left and sector on the right (N=66) ........ 14
Figure 2. Country distribution .............................................................................................................. 15
Figure 3. Company size ....................................................................................................................... 15
Figure 4. Industry category .................................................................................................................. 16
Figure 5. Data value chain ................................................................................................................... 19
Figure 6. Levels of data from the platform business user perspective ..................................................... 19
Figure 7. Data sources and creation of the value of data ....................................................................... 27
Figure 8. Amazon Retail Analytics: Basic vs Premium package ............................................................... 30
Figure 9. My business can easily access data collected by the platform that is important to my business .... 47
Figure 10. Do you have access to the following types of data held by the Main online platform? .......... 47
Figure 11. How much processing does the Main platform carry out on data it provides to your business? ...... 48
Figure 12. Needed but inaccessible data held by the online platform .................................................... 48
Figure 13. Difficulties experienced in accessing data from the platform ............................................... 49
Figure 14. Business user survey: Does your business analyse raw data provided by the platform, by third-party sources (data brokers) or collected by itself? .................................................. 50
Figure 15. Data collection by businesses themselves, by data type ..................................................... 51
Figure 16. Use of third-party sources to access any types of data .......................................................... 52

LIST OF TABLES

Table 1. Amazon Brand Analytics, as accessed by a seller ...................................................................... 31
Table 2. Types of data and analytics accessible by Amazon Marketplace sellers ...................................... 32
Table 3. Types of data and analytics accessible by eBay sellers ............................................................... 36
Table 4. Google Play Console: key metrics for app developers ................................................................. 37
Table 5. Types of data and analytics accessible by Google Play app developers on the platform ............... 39
Table 6. Analytics for Booking.com business users ................................................................................. 40
Table 7. Types of data and analytics accessible by properties listed on Booking.com ............................... 42
Table 8. Cooperative platform as a way to generate data ......................................................................... 51
Table 9. Data brokers’ ownership status and locations ............................................................................. 56
Table 10. Data types collected and supplied by data brokers .................................................................... 60
Table 11. Data collection sources and methods of analysed data brokers .................................................. 68
Table 12. Analysed data companies’ compliance with GDPR .................................................................... 72

LIST OF BOXES

Box 1. Key definitions of technical data collection methods ..................................................................... 65
Box 2. Issues with data protection and privacy .......................................................................................... 71
Executive summary

Setting the context

Online platforms create value by using data to facilitate interactions (for example, commercial transactions) between users. This means that data is at the core of the platforms’ business model and they use it to provide and improve their services. When a platform acts as a marketplace to business users, some data sharing will be part of the service offered. Business users can draw on such data in order to reach, understand, and maintain their clients, although the scope and conditions of data sharing may also be subject to disputes between platforms and business users.

Platforms collect and process a variety of data that can be volunteered or observed, individual-level or aggregated, personal or non-personal. The data in their possession allows the platforms to understand the preferences of customers and their reactions to market signals, including changes in prices and product characteristics. This puts online platforms in a unique position as they are able to observe the functioning of the market in real time. Data is thus a key source of market power. In other words, platforms’ decisions on what data to share, with whom, and under which conditions have far-reaching consequences to all the participants in the market.

When taking decisions on data sharing, platforms perform a balancing act between several competing pressures. From the platforms’ perspective, they have an interest in facilitating as many transactions as possible, which is an argument for sharing enough data so that businesses could attract their customers better. Yet providing individual-level or personally identifiable information to business users may undermine a platforms’ business model as business users may draw on this data to bypass the platform in the future. Platforms may also use data to ensure their own competitive advantage in the market vis-à-vis other platforms or, in the case of vertically integrated platforms, in relation to their own business users. Finally, collection and sharing of personal information may conflict with privacy concerns and has therefore become subject to increasingly thorough regulation by the public sector (e.g. GDPR), which means that platforms have a legal obligation not to collect or share some data, or should ask for explicit consent from data subjects before collecting or sharing data.

From the perspective of business users, it would be useful for them to receive data about the competition in the market, in particular about competing businesses, products or services. Yet, they would also not want that detailed information about them was shared with their competitors. Again, platforms must balance these needs by allowing the business users to understand the market without breaching their trust. Given that business users themselves are very different in terms of their size, market power and analytical capacities, their data needs are also very different: while some ask for raw, granular data, others can only make use of highly-processed data and prefer to get actionable data-based insights.

---

1 See Chapter 3 of the paper where we discuss the types of data
3 Any data sharing decision is indeed a private decision of a platform as only few areas benefit from rules on mandatory data sharing or industry codes of conduct, such as automotive, financial services, agriculture.
Sources and methodology

The paper draws on extensive desk research, business user survey and interviews. The desk research focussed on an overview of the research literature (both grey and academic) as well as various news articles, blogs and forums. Further, we carried out over 60 interviews with online platforms, their business users, data brokers and non-governmental organisations representing the key stakeholders. The business user survey was conducted in November-December 2019; it covered nine EU countries, with the sample size of 1,667. Given that the data needs and sharing practices are sector and platform-specific, we focussed our research on three sectors: e-commerce, dissemination of apps and software, and accommodation/hospitality. For the analysis of data companies, we analysed 15 companies, serving the data needs of the three sectors investigated in this paper.

Business users: data needs and access to data

In the paper we identified three general dimensions of data relevant to platform business users:

- the type of data by object (customers, businesses, user behaviour, markets, transactions, etc.);
- whether the data relates to an individual business, or to other businesses (competitors) on the platform / whole marketplace.
- by the level of data processing and its value, from raw datasets to insights guiding business decision-making.

The paper shows that the kinds of data provided or not provided by the platforms (Amazon, eBay, Google Play and Booking.com) are rather similar. Access to data as well as advanced analytics are granted to the extent that it could generate more income for the platform as well as for the business users. In such a case the key question is whether the business users can take full advantage of the data provided to them. Further, a significant share of businesses signal that they are experiencing data access problems. This was very visible in interviews where businesses, especially the bigger or stronger ones, felt strongly about the data access. Their key concern was getting access to data so that they could use it to innovate and keep up with the competition. A recurring issue was also the power of the vertically integrated platforms and especially the extent to which such platforms may use data to develop their own competing products. In the next two sections we briefly explore, firstly, the findings concerning taking advantage of the available data and, secondly, the evidence concerning data that is not shared with the business users.

Making use of the available data

In order to understand to what extent business users take advantage of data that is available to them, we reviewed the types of data collected by the platforms. The platforms collect data on customers, businesses, user behaviour, markets and transactions, and other dimensions. When asked, the business users usually report that all such data is of interest to them. Most of the businesses use (or would like to use) both, data on their individual performance, as well as data on their competitors and overall market trends. Nevertheless, our study also shows that business users tend to think about data in very different ways. Further, the interviews with, for example, Amazon sellers showed that businesses do not always know what data is collected by platforms and what is accessible and useful to them.

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

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

Data that is not available to business users

Our research showed that a significant share of business users express dissatisfaction with regard to the level of data access provided to them by online platforms. The business user survey showed that access to data possessed by online platforms is of concern to around a third of surveyed business users who reported that they cannot access at least some data that is essential to their business. Generally, although the platforms collect and analyse loads of data, only a fraction of this is provided to other players. The platforms do not share the raw big data on day-to-day activities, as well as detailed data on customers and competitors.

Our paper identified three groups of concerns that business users express with regard to data sharing.

The first is related to lack of access to personal data, such as customers’ e-mail address. Some business users, especially in the hospitality and e-commerce sectors consider such data of key importance to them so that they could establish a more direct client relationship. Other personal data collected by platforms, but usually not provided to business users include, for example: telephone, address, credit-card data. As confirmed by our desk research and interviews with the platforms themselves, this data is not provided for a number of reasons. Firstly, this is not considered compatible with the platform’s business model as business users may use direct communication to bypass platforms in the future. Secondly, platforms consider that a consistent client relationship and data protection is part of the client experience. They are wary that direct access to the clients by business users may result in a surge of unwanted marketing messages (this argument was not supported by the interviewed business users). Finally, the personal data protection regulation (including GDPR) puts obligations on platforms in terms of data sharing and management, including the obligation of getting explicit consent from consumers to collect and share their data. According to the business user survey, legislative or regulatory restrictions is indeed acknowledged as the key reason for not getting access to data.

Secondly, business users need data that help them to stay competitive, innovate and develop their products and services. Partly this is related to data on, for example, search keywords, search volumes, consumer behaviour in reaction to different price signals. Platforms do share such data to a certain extent (e.g. three most important key words), however some interviewed business users felt that this does not give them sufficient level of detail. Partly, this is also related to data about competitors and their products and services. In this case however, both the interviewed platforms as well as business users expressed
understanding that the level of detail is naturally limited as businesses would not want their individual business performance information to be made available to others.

The third concern is that platforms are taking advantage of data to promote their own products that are very similar to those offered by their business users. This is primarily pertinent to vertically integrated platforms with significant market power. So, in the business user survey, 58% of respondents reported that the platform itself offers the same (or very similar) goods or services to those that their businesses offer on the platform. Among these respondents, 55% argued that online platforms are favouring their own goods or services vis-à-vis the same (or very similar) goods or services offered by their businesses. Unique and comprehensive datasets on all the firms and their consumers operating in the marketplace can give a huge business advantage to the platform operators. The key ways of favouring include ranking, placement of advertisement, pricing and other – all of these are enabled by the data collected by platforms. Some interviewed business users argued that platforms (specifically – Amazon) are using data to monitor which goods have the best margins in the market and then move into offering such goods themselves. The interviewees did not want to present specific examples, in fear of possible repercussions. However, media articles provide stories on unfair competition between Amazon private labels and third-party Marketplace sellers. Amazon’s practices of sharing seller data between Retail and Marketplace, as well as sophisticated use of metrics to boost its own sales were also corroborated by the platform’s former employees.

The role of data companies

The businesses that need more data than they can get from platforms pursue two broad strategies: (1) collect and analyse data themselves, from sources available to them; (2) rely on third party providers (data brokers). Four-fifths of the business user survey respondents (81%) indicated that they collect some data themselves; the most prevalent data type is identification details of own customers (57% of respondents collect this data), followed by business performance data (55%) and analysis of market trends/ developments (55%). Further, a third of the surveyed companies (33%) reported that they use third-party sources (data brokers). Most of the interviewees - especially in the e-commerce sector - reported that they use the services of third-party data and analytics providers.

Companies specialising in data and insights fill important data gaps, especially with regard to data on competition and actionable business insights. According to PwC estimations, data companies earned $21 billion in 2018. Some data brokers specialise to cover specific sectors, such as applications (App Annie, App figures, Apptopia, Mobile Action, Sensor Tower), e-commerce (Algopix, Jungle Scout, Sellics, Teikametrics, Terapeak) or hospitality (AirDNA, Beyond Pricing, Uplisting, Wheelhouse, Skift). Other data brokers, such as Similar Web and Zirra provide data on multiple sectors.

Data brokers use highly advanced technical methods to extract data, or they buy data from online and offline sources. A lot of data is scraped from the platforms. According to a blogger and speaker on Amazon, the platform’s approach to the data companies is “to let them ride” by leaving “the backdoor open”. Yet Amazon also banned several entities that were acquiring user data from the platform in

---


violation of its privacy policies. Some other platforms do not permit scraping for commercial purposes, but that does not always prevent data companies from collecting data this way.

Another key source is crowdsourcing business user account data. Some third-party data providers ask online sellers to share their marketplace information, and then link the data of thousands of users to draw market insights. For example, Jungle Scout collects data from a large number of sellers (over 225,000) who have opted in to share their sales information. When merged with the data gathered by scraping the platform’s front-end (e.g., Best Seller rank on Amazon), this can yield quite precise estimations and extrapolations. Similarly, if AirDNA users wish to receive performance analytics, they will be asked to upload their Airbnb host IDs. After doing this, they can see their performance trends, comparative and financial analysis on all vacation rental listings.

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

For instance, as explained by several interviewed Amazon sellers, Jungle Scout and other providers, such as Helium10, AMZScout and Unicorn Smasher, supply them with comprehensive market insights and competitor overviews. Obviously, these data brokers provide estimations based on what data they could gather rather than exact information. Nevertheless, the estimations are said to be “spookily accurate”. Similarly, data providers for app developers, such as AppAnnie and SensorTower, offer comprehensive app market data, including performance of specific apps and markets. Interviewed app developers mentioned that they use the sources together with the app store data extensively. In the accommodation/hospitality sector, companies such as AirDNA provide insights based on data that the OTAs do not share. For example, in late 2015 Airbnb stopped providing the overall real-time reservation data. AirDNA, in turn, uses an algorithm based on 16 indicators picked up in historical data to determine the reservation status for each listing. They argue that their algorithm has an error margin of only 5%.

The data companies’ market is very dynamic and fast-paced. This paper identified a number of issues, illustrating the key challenges and limitations of data brokers. Firstly, the data companies remain highly dependent on data sharing policies of platforms. For example, Amazon until recently provided exact and broad match search volume and product relevance data via one of its APIs. It was feeding several third-party software providers such as Viral Launch and Helium10 until late 2018, when the platform removed these metrics from the API. Another platform, Allegro made significant investment to develop new data products (Allegro Statistics) that are now provided to its sellers; this is endangering the business model of third-party analytics providers.

---

Secondly, the data needs of platform business users are often very specific and concern platforms that they use. Such data cannot be easily scraped or estimated by the third-party data providers. It includes information on real-time activities on the platform (e.g., X currently has product Y added to the shopping cart), which would allow to effectively address the customer; transaction-related data about the customers, sales activities and listings of specific business user.

Thirdly, the huge amounts of data that data brokers collect, store, possibly re-personalise and disseminate are of interest from the regulatory perspective, first and foremost due to privacy concerns. Most individuals or companies are unaware of what information data brokers collect on them or even that they collect information at all. Due to this asymmetry, the data broker industry has been often characterised as opaque, non-transparent, arbitrary, biased, unfair and unaccountable. Interviewees from the data brokers argued that they are taking actions to make sure they are compliant with data protection and privacy laws, such as the GDPR. However, other sources show that such compliance has not always been properly ensured. For example, a few months after the GDPR came into force, Privacy International filed a complaint against seven data brokers: Acxiom, Oracle, Criteo, Quantcast, Tapad, Equifax, and Experian. The main argument was their failure to comply with data protection principles (such as acquiring consent, providing detailed and transparent information for the data subject access requests) and exploitation of data in unknown ways.

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

Reflexions on mandatory data sharing by platforms

The question whether there should be horizontal obligations for platforms to share more data is complicated and our research did not lead to unanimous results. Firstly, providing everyone with the same access to data may be detrimental to the business users themselves, whose financial, client and performance information could be disclosed to competitors. Secondly, the impact of data sharing should also be assessed from a consumer perspective. Platforms invest a lot in customer satisfaction, and data protection is part of the consumer experience. Consumers prefer to deal with centralised services rather than with multiple actors asking for consent, and they are not willing to be flooded with marketing messages. Finally, mandatory data sharing may have a negative impact on platforms’ business and profits and might act as a disincentive for innovation.

Further research

This analytical paper presents an overview on the state of the art of data access for platforms’ business users. It also raises new questions for further research. The analysis revealed concrete issues in specific

---

segments of the platform economy. Firstly, the most problematic cases concern vertically integrated platforms with significant market power, who can use data to their advantage, while not sharing it with their own competing business users. Secondly, this paper looked into three sectors of the platform economy: e-commerce, app development and accommodation/hospitality. In order to get further insights on data access, further investigation into other sectors not covered in this paper is needed. Thirdly, further research would be useful into the legal and technical aspects of the platform data and sharing, including interoperability.
1. Introduction

Due to technological innovations and penetration of technologies into new areas, many entities from different industries have the tools to collect extensive and diverse types of data on products, processes, customers and internet users more generally. Some sources estimate that as of 2019, the nearly 4.5 billion Internet users generate about 2.5 exabytes\(^{16}\) of data each day\(^{17}\). A very significant share of this enormous amount of data comes namely from online platforms, that collect data on all the activities taking place on them, both actively and passively. To illustrate, as of January 2019, out of 20 most visited websites on internet, 11 were what we classify as online platforms\(^{18}\), also taking four top places (Google, YouTube, Facebook and Baidu)\(^{19}\). Google alone processes over 3.5 billion searches\(^{20}\) and 20 petabytes of data every day\(^{21}\). Another four petabytes of data are created on Facebook daily\(^{22}\).

Together with the increase in the amounts of data collected, its importance for business success has also increased dramatically in recent decades. In many cases, data is no longer the mere accessory of a process or product, but often an integral part of it. It has also become a key commodity of many market players interested to improve their services and products, expand the market, understand competition, and apply many other data-driven solutions for their business growth. The existing body of research emphasises data as a key input for optimisation of business processes and innovation\(^{23}\). Some studies suggest that companies using data-driven innovation benefit, on average, from 5% to 10% faster productivity growth than similar companies that do not apply data-driven innovation\(^{24}\). Basically, having the most accurate and comprehensive dataset ensures market power\(^{25}\).

From the economic perspective, as a commodity, data has two important characteristics. First, it is non-rivalry in consumption. This means that the same data record can be used by different actors for different purposes at the same time, because it is easily reproduced and distributed almost free of charge. This also means that, theoretically, different companies and individuals can exploit the same data, thus providing opportunities in markets for more participants, and allowing for dynamic entrepreneurial activity. This characteristic distinguishes data from many other assets and tangible goods, which are available in limited quantities and at a significant cost.

However, while the non-rivalry of data in consumption implies that it is not a scarce resource, the excludability characteristic of data works in the opposite direction. That means that the owner of a dataset may exclude other users from consuming by denying them access. In this way data can be

\[^{16}\] 1 EB = 1,000\(^6\) bytes = 1,000,000,000,000,000,000 B = 1,000 petabytes = 1 million terabytes = 1 billion gigabytes.


\[^{18}\] Similarweb & Alexa.


monetized, and collecting it becomes economically attractive. In fact, the recent dynamics in the data markets rather show that companies - platforms and their business users alike - are striving to acquire a 'data advantage' over rivals.

Currently, as illustrated above, most of the global data is collected and processed by several large players. Mediating millions of transactions every day, the largest online platforms have access to the data of all their user groups. They collect and aggregate vast amounts of data in their day-to-day operations, allowing them to have a comprehensive market overview that is more valuable than separate data on individual transactions. This leads to a substantial data asymmetry between firms and the platform which they use to market their products. Such massive amounts of data are considered to place the platforms in a position of almost unassailable dominance among economic actors.

Most business users of online platforms heavily rely on data and analytics. These have become a key tool for the entities conducting business on online platforms to determine which product, price and advertising is best to maximise their profits, understand their customers and sales, and manage the supply chain efficiently. All platforms provide them with at least some access to such information: granting access to certain types of data spurs the development of new products, services or complementary applications that enhance the value of the platform itself. However, many earlier investigations show that online platforms quite often refuse to share some of the crucial data with their business users, preventing these firms from providing products or services that require this data as an input.

At the same time, the largest online platforms become the focus of many third-party data providers, offering the platform business users an alternative source of data and insights. The growth of online platforms and numbers of internet players more generally goes hand in hand with data accumulation by third parties, who use these data for analytics and value creation. The importance of data companies also seems to be growing rapidly. European Commission forecasts that the overall data market in Europe could be worth as much as €106.8 billion by 2020. Similarly, PwC expects that the global data economy will be worth more than USD 400 billion by 2025. Some authors even argue that most of the data is actively traded and accessible to anyone willing to pay for it, indicating a healthy data market. Often

---


---
many substitutable sources exist to acquire data, and having large in-house datasets does not necessarily confer an entrenched advantage.

Nonetheless, there is little knowledge whether such tradable data sources, third-party analytics providers or data brokers can indeed provide businesses with sufficient information to overcome the data disadvantages in the platform economy. Moreover, so far, the discussion about the data needs of business users has been rather vague. In this context, the aim of this paper is to investigate and provide comprehensive insights on:

- The general data needs of platform business users, including a data typology, allowing to specify the discussion on platform business user data demand.
- Platform data collection and sharing practices, the current limitations in accessing data collected and held by platforms, and rationale for not sharing the data, considering potential trade-offs between the interests of business users, consumers and those of the online intermediaries.
- The ability of business users to collect data relevant for their businesses from alternative data sources, allowing to meet their data demand, limit the dependency on the online platforms, or triangulate data sources and build more valuable insights.

Exploration of these topics will address several knowledge gaps and research questions, identified by the Expert group in their report on data in platform economy, including:

- Review of the actual data sharing practices in a range of online platforms through case studies.
- Review of sector-specific data needs by platform business users, and purposes of data use.
- Analysis of how businesses and 3rd parties obtain and exploit data, considering elements of digital skills and contractual constraints.
- Research into the role of data brokers in the ecosystem for facilitating data sharing and making its benefits available to players of all sizes.

The following Chapter 2 of this paper presents the scope of this analysis and methodology of data collection applied in answering these questions. The third chapter starts the analysis with an overview of the main data types and characteristics that platform business users need to make informed business decisions as well as successfully compete and scale their operations. The fourth chapter reviews the data collection and sharing practices of four selected platforms from different sectors of the platform economy, vertically integrated and non-integrated. After identifying the gaps in data access, Chapter 5 focuses on the availability of alternative sources for platform business users to acquire data. Finally, Chapter 6 looks in-depth into actors providing alternatives to platform data - third-party data brokers. Finally, Chapter 7 provides a summary of the main findings as well as directions for future research.
2. Scope and methodology

2.1. Scope of the analytical paper

To investigate the business user data needs, this analytical paper focuses on three broad sectors with the platform economy: e-commerce, dissemination of apps and software, and accommodation/hospitality. These were selected as the sectors and economic activities most and increasingly reliant on online platform intermediation services. To illustrate, between 2014 and 2018 online hotel distribution in Europe grew by 46.7%\(^{38}\), while the growing e-commerce sector in Europe was estimated to reach sales of 621 billion euros in 2019\(^{39}\). Meanwhile, only a miniscule share of mobile application distribution takes place outside app stores. The following parts of the paper, therefore, looks into the platforms, business users and third-party data providers operating in these three sectors.

For the in-depth analysis of the platform data collection and sharing practices, we selected four platforms to represent each sector. The specific platforms were selected as the most used ones globally and in Europe: Amazon and eBay for e-commerce; Google Play app store for app developers, and Booking.com for hospitality sector. This selection allowed us to limit the scope of our focus and get a quite comprehensive view of the broader data-sharing practices in the platform economy, because:

— The selection covers platforms that intermediate commercial transactions between businesses and their customers. This multi-sided nature allows them generate loads of data on these activities, which becomes important for business users to further develop their business models, strategies and products.

— VVA study\(^{40}\) indicated that these three groups of business users have quite different data needs and levels of access to data. The selection, therefore, will allow us to cover the existing variety and look into the issue comprehensively.

— The platform selection allows comparisons between vertically integrated and non-integrated platforms. More specifically, most existing major app stores are vertically integrated (i.e., they offer own apps on their stores), while most OTAs are not (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 two marketplaces to cover both models.

— These are the largest/ most successful platforms, covering most business users. Many interviewees agree that business “must” be on them to be successful in their respective businesses. For example, (although e-commerce is the most diverse area of the three selected) Amazon is the major global and EU player, with a wide variety of products, as well as attempts to copy Amazon model by other platforms. Google Play is the largest app store in Europe and


globally, with 3.3 million apps in 2019. Meanwhile, Booking.com has 66% OTA market share in Europe in 2017⁴¹, with most European accommodation properties quite dependent on it⁴².

Further, to investigate the alternatives for business user data access and third-party data brokers more specifically, we broadly overview a selection of data companies, servicing the businesses from the three investigated sectors (app development, e-commerce and hospitality).

2.2. Methodology of data collection and analysis

Generally, the data and evidence used in this analytical paper was collected through desk research, interviews and an online business user survey. Given that most of the data on the research questions was collected from actors involved directly in the platform economy and representing particular interests, the research team triangulated and fact-checked the information feeding into this report.

Desk research included the overview of the research literature (both grey and academic) on the issues of data in the platform economy. Given the existing literature focusing on the specific questions of this analytical paper was rather scarce, the research team also reviewed various news articles, blogs and forums, some of them also serving as primary data sources. The researchers used this information for two purposes: as a source 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, platform’s own presentation of their policies and interview data). The full list of reviewed sources is provided at the end of this report.

Furthermore, the analysis presented in this paper is based on over 60 interviews with online platforms, their business users and associations representing each of these groups. The lists of main interview topics and interviewees is provided in at the end of this document. The figure below shows a breakdown of the total completed interviews, including all stakeholders, according to their types (business users, online platforms, and data brokers) and sectors (e-commerce, accommodation/hospitality, apps and software, or multiple sectors).

FIGURE 1. TOTAL COMPLETED INTERVIEWS ACCORDING TO THEIR TYPE ON THE LEFT AND SECTOR ON THE RIGHT (N=66)

The interview transcripts were coded using NVivo software for qualitative data analysis, using codes corresponding to the main research questions of this paper. This allowed us to employ all the evidence from interviews in a comprehensive and systematic manner.

---

⁴¹ Statista.
Finally, the analytical paper also presents the results of a quantitative **online business user survey**. The survey was administered using SurveyGizmo survey software and disseminated to respondents of commercial online B2B panels. The final sample size of the business user survey is 1,667 responses from nine EU Member States (see the figure below).

**FIGURE 2. COUNTRY DISTRIBUTION**

Source: PPMI; n=1,667

The majority of the sample (55%) is comprised of micro and small companies (with less than 50 employees). Large enterprises account for 19% of the sample, and medium – 26% (see Figure 3).

**FIGURE 3. COMPANY SIZE**

Source: PPMI; n=1,667

In terms of sectorial distribution, a quarter of enterprises in the survey sample are from wholesale and retail trade, a fifth from industry (manufacturing and construction), see Figure 4 below. It must be noted that the business user survey targeted the business users of all internet platforms that fit the study definition, not only e-commerce sellers, accommodation providers and app developers. More specifically, the enterprises that participated in the survey use platform for one or several of the following purposes: selling goods, selling services, distributing apps or software, listing accommodation and advertising. Therefore, the survey results reflect a broader overview of data sharing in P2B relationships.

---

A detailed methodological report of the survey is presented in a separate document, attached to the First Interim Report.
The synthetic analysis of all the collected data, focusing of the main research questions of this paper, is presented throughout the paper. The interpretation of these results should take into account the limited scope of this paper, not allowing to generalise them to the overall platform economy or all platforms in the specific analysed segment. In addition to this, the researchers did not have the same level of access to platform and data broker data as their registered or paying users do. Therefore, we had to rely on the information on the data access as it was reported by the interviewees or collected from secondary internet sources.
3. Platform business user data needs

This section focuses on the data needs from the platform business users’ perspective. We start the analysis on the assumption that the need for data by platform business users is driven by the goal to succeed and maximise profits from activities conducted on the platforms. Data (of various types and levels of processing), therefore, should be available and used instrumentally to achieve this goal. We first provide an overview of the most relevant classifications of data in the view of platform business user needs, as well as propose a way to approach data for the further analysis. Second, we review the sector-specific platform business user needs of data. We end the chapter with a summary of insights into the existing evidence of business data needs and the limitations to acquire this data.

3.1. Types of data for platform business users

Many ways to classify data exist, and they often differ very significantly content-wise and in the choice of dimensions. For example, a report by DG Competition provides several typologies by source (volunteered, observed, inferred) and form (individual level, bundled (anonymous) individual-level, aggregated-level, contextual) of data, as well as its personal (which can be further distinguished between volunteered, observed and inferred44) or non-personal45 character. Meanwhile, OECD46 in B2B data flows distinguishes between global value chain (GVC) data, engineering, IoT and financial/human resources data. The desk research showed that in the broader platform data discussions online data is also sometimes classified (implicitly or explicitly) by:

- Source (e.g., web, IoT, surveys, etc.);
- Level of structure, i.e., structured (including some kind of identification allowing to organise the datasets accordingly (e.g. customers and links it to their behaviour), or not (e.g. likes, tweets, clicks and videos and so on);
- Level of specificity (platform, sector, general);
- Object (businesses, countries, individuals, etc.);
- Level of aggregation and granularity.

This variety illustrates to some extent the lack of common vocabulary in operationalising the rather vague concept of “data” and its needs in the platform-to-business relationships. Nonetheless, 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 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:47

- 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”).

The conducted interviews and desk research largely allowed to validate this typology and operationalise the categories, clearly distinguishing one from another. However, the collected evidence showed that several additional dimensions of data should be taken into account concerning business user data needs, data access and lack thereof.

First, especially in relation to the four latter data categories (i.e. data on individual transactions, business performance, user behaviour and market trends) and their usefulness for businesses, the interviewed business users emphasised the distinction between:

- Data on their 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. This data, as we demonstrate in the following chapter, is extensively provided by platforms to their business users.

- Data on the competitor or overall listings, customers, transactions and performance of businesses on the platform. 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. However, as our platform case studies show (Chapter 4), this data is shared with the business users to a very limited extent. Meanwhile, studying competition is one of the key steps for businesses to succeed on platforms.

As we discuss in the following sections, 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. Therefore, most discussions about the lack of platform data access revolve namely around this information about the market trends, competition and comparative business or product performance, market niche identification and so on - accessible by platforms but not by other market players.
Second, several conceptual levels of data and its use and applications exist, related to the data value chain\(^{49}\) (see the figure below). Based on the data value chain framework, data follows a consequential approach, and at each step its different data features 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 a its 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 (we elaborate on this in Chapter 4).

**FIGURE 5. DATA VALUE CHAIN**


From the business user perspective too, data is as useful, monetizable or needed, as it can be effectively used in making business decisions. Analytically, therefore, what is very broadly referred to as data, can be divided into several levels, as seen by data scientists, closely related to data aggregation and processing, and data-based decision-making (see the figure below). First, it is raw data collected by the platforms, business users or third parties. It is used as an input for specific actions or in analytics, outputs of which is the second level. The third level is insights (or the actionable guidelines for business decisions), developed based on the analytics, and directly applicable in business processes.

**FIGURE 6. LEVELS OF DATA FROM THE PLATFORM BUSINESS USER PERSPECTIVE**

Source: developed by the authors.

Not all data collected by players in the digital platform economy is equally relevant for business users. Acquiring and using each level of relevant data, as presented above, is related to specific conditions and prerequisites.

Raw data

As illustrated in the figure above, it is useful to distinguish between two kinds of raw data: one that is used for further business analytics (i.e. the big data collected by platforms and other players); and data which is needed to perform an action (from contacting clients to developing data products).

To be used for further analytics and then business decisions, raw data used for further analysis needs to be collected continuously and have certain characteristics. The big data frameworks talk about four Vs of big data, determining its value: volume, variety, veracity (i.e. trustworthiness) and velocity (i.e. frequency of incoming data)\(^5\). Conditions to implement these criteria should be ensured before starting the data collection. Although some improvements can in data quality can be made once it is collected, it mostly depends on the steps taken at the starting point. Online platforms, as explained, are usually the players with the greatest capacities to collect and store large volumes of real-time data on every activity that is taking place on their websites or apps. Relatively few platform business users from selected sectors have the capacities to set up such data collection operations themselves.

Furthermore, the content and character of raw data collection differs - and sometimes to a great extent - by sector in which business users operate. For example, app developers have the possibility to gather large amounts of real-time data on how their products are used directly through their applications, while many hospitality sector players still use paper-based forms and rely on customer willingness to collect data on them and their satisfaction.

Generally, it seems that rather few business users are interested in getting such raw and unprocessed data, which could be used in further analysis, from platforms directly. The main reason is that they do not have enough capacity to derive value from unprocessed data themselves (as illustrated in Figure 5, raw data has least immediate value). We elaborate more on the business user analytical capacities in Section 3.2 and Chapter 5, and on actors helping businesses to fill the capacities gap in Chapter 6.

Then, access to specific kinds of unaggregated data in some cases is also a prerequisite for the implementation business strategies or development of data-based products. For example, customer contacts (which could be classified as raw data) are essential for business users to carry out direct marketing campaigns. Another obvious example for the application of raw data is data products (e.g. applications and software, business analytics, predictive maintenance of a machine) and AI models used in business processes. For instance, a study by Accenture estimates that appropriate use of AI could achieve additional annual growth of 1.9% in retail and wholesale sectors, and 2.3% in manufacturing in mature economies\(^5\). Nonetheless, to programme and train AI, as well as to use it later, a lot of quality data is necessary as an input.

In summary, the business user needs for unaggregated data mostly concern specifically the data as a prerequisite to act (e.g., distribute a marketing campaign or develop a data product), and not raw data to use for analytics. Most business users, especially in e-commerce and travel sectors, usually do not have the capacities (nor the need) to work with unaggregated and unprocessed data, generated by platforms.


Analytics

The essential value of data lies not in collection and ownership per se, but in the information which can be used to improve competitiveness of products/services or to develop new products and services. Therefore, the value of data for businesses increases with its level of processing. The main prerequisites for generating analytics are the competence and capacity (either in-house or external) to analyse raw data. This includes appropriate hardware and software, and, more importantly, employees with the necessary skills. Specific uses of analytics then depend a lot on the business model of platform users, type of a product, existence of possibilities to increase efficiency and so on.

The evidence collected in preparation of this paper shows that business users usually lack analytical capacities to derive such value from data. Most platforms, in turn, often provide some analytics data. Nonetheless, the level or detail and granularity of this data is often very low. In addition to this, the analytics provided mostly concern the performance of individual businesses (i.e., business get the analytics about themselves), while the business users also express the need of a broader overview of the processes, activities and competition taking place on the platforms.

Insights

The most usable and monetizable level of information for all types of business users are insights and actionable guidelines, stemming from the results of analytics. While analytics can be understood as an input into the evidence-based decision-making process, specific insights and action guidelines are the ultimate output (e.g., optimise product titles, based on the search behaviour data; develop a new product based on predictions what the consumer will buy; add tea kettles to hotel rooms to attract Asian tourists, based on customer feedback data; change the monetisation model of an app based on user engagement trends; etc.). Such insights can concern various aspects of running of business. The interviews and desk research most often mention the following examples:

- How to improve business processes,
- How to improve or build new business models,
- How to build a better customer experience,
- How to increase customer satisfaction and conversion rates,
- Whom to target in marketing, ads and recommendations,
- How to adjust pricing,
- How to customise products or services
- What new products or services to develop,
- How to improve listings, and so on.

This level of data-based insight is quite often provided by the platforms, especially for new business users. Various platforms provide detailed guidelines, “tips”, seasonal advice and similar, based on the analytics that they generate. At the same time, many businesses report arriving to such insight themselves with the analytics provided by platforms or acquired from other sources.

3.2. Sector-specific data needs

Businesses from different sectors face different challenges in their day to day operations, so their needs vary a lot depending on their activities and platforms used. For example, earlier research shows that hospitality platform users highly value customer identification data (which is necessary to contact customers for promotional activities), whereas user behaviour data (which is necessary to target communication to individual customers and personalise) is of particular value for app stores and e-
commerce platform users\textsuperscript{52}. These insights were strongly corroborated in the interviews with businesses conducted in preparation of this analytical paper.

Furthermore, a VVA study\textsuperscript{53} also shows a great difference in the data savviness of different size and type of businesses – from large hotel chains, to very small hotel owners, from app developers to small shops selling products on market places, or from large corporations monitoring and optimising their presence on search engines, to minor website owners. Some types of users, such as app developers, have great opportunities to collect data about the use of their products themselves (sometimes even balancing on the narrow line of serious privacy breaches). In contrast, because of the nature of their offerings, retailers and hotels do not have such possibilities to monitor all aspects of how their offerings are purchased and used.

Related to this, the analytical capabilities vary a lot by sector of the business user and often their company size. While some business users would be very happy with ‘raw’ data, others would prefer ready and actionable insights. In fact, some earlier analyses as well as interviews with business associations show that data analysis and required analytical capabilities are often the “pain points” that need to be addressed. For example, in a 2015 survey, almost a third (32\%) of travel and hospitality industry marketers rated data science as the skill they most lack, with programming coming in a close second at 28\%.\textsuperscript{54} Although comparable data was not available for e-commerce seller and app developers, it is obvious that the latter are considerably more data-savvy.

Further we provide an overview of more specific data needs of businesses operating in different industries.

3.2.1. E-commerce sellers

Interviews with e-commerce sellers and desk research data shows that the distinction between the data on the seller’s own activities and overall activities on the platform, as explained above, is very relevant. Regarding the first category, the sellers expressed a need for the following data:

- Customer data, including phone number and address that are crucial to safely deliver the purchased goods.
- 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.
- Customer interaction with their listings (e.g., how do customers reach their listing pages, click-through rates, bounce rates, conversion, reviewer characteristics, etc.).

Most (but not all\textsuperscript{55}) of these indicators are provided by the platforms for their business users, although not always at the level of actionable insight.

What sellers commonly identified as an unmet data need is the market analytics allowing them to better understand customer behaviour, market trends and competition (i.e., the data on overall platform


\textsuperscript{55} E.g., the email address are provided very rarely, phone numbers are also excluded by some platforms (e.g., Etsy).
activities). According to an article by Akter and Fosso Wamba, analytics based on such data can be applied in e-commerce for five main broad purposes: needs identification, market segmentation, decision making and performance improvement, new product/ market/ business model innovations and creating infrastructure and transparency. The interviewed marketplace sellers mostly emphasised performance improvement and product/ business model innovations as the main purposes for which they would like to access more data generated by platforms. However, these are difficult to achieve with data on their own business only.

More specifically, to be successful in online marketplaces, sellers need several key 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 ranks, and so on. This includes competitor’s sales data. If they had it, sellers could 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. Competitor research could also help to uncover hidden products with large sales potential. For large companies, and especially luxury brands, sales data and details of counterfeit products is also important.

- **Customer behaviour**, especially search keywords and search volume data, to optimize their stores on platforms for search, as well as identify new products to sell. Understanding their reviews and shopping habits, responses to texts, images and videos is also crucial.

- **Price data** of products similar to theirs, to develop (dynamic) pricing strategies. For example, Amazon’s algorithm would choose which seller would win 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.

The data collected for this paper corroborated the findings of previous research, 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. Our business user survey shows that 52% of enterprises whose key activity on platforms is selling goods, 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

---

60 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.
e-commerce platforms. Without data access, third-part sellers cannot viably compete with the marketplace controller which has unique insight into the market.

Interestingly, several interviewed e-commerce sellers, when asked about what data is most important to successfully run a business, mentioned the data, metrics and indicators that are already provided by platforms. One of them explained that they are not aware of all the metrics and indicators that can be useful to them, if the platforms or other providers do not provide them. He provided an example: during a phone call with an e-commerce platform, the platform representative mentioned that he should do something about the issue that most of his customers find his listings on the platform (internally), and not through external links (e.g., advertising). He was surprised because he was not only not aware about this issue; he also did not know that such an indicator is worth tracking for business decisions. The platform, meanwhile, tracks this information, but does not provide the sellers with access to it.

### 3.2.2. App developers

App developers are a specific group of business users, as data is not only needed to understand their business performance - it is the key component of the products they offer. The specific datasets that are needed (e.g., geolocation, speech recognition, etc.) depend a lot on the type and specific characteristics of an app.

App developers also have broader possibilities to collect data themselves using their own software. Apps can interact with the user’s device so as to retrieve and share such information as photos, biometric data, financial information, contact details, login credentials, correspondence, location, IP number, etc. They can also activate some of the device’s functions, such as the camera and microphone, that can be used to collect additional personal data. In fact, interviewed developer representatives confirm that this is the most important data for their activities. On the negative side, however, the extent of their data collection activities sometimes even clash with the data protection rules. This business user group can also be characterised by greatest capacities to run analytics on this data and make insights on how to further develop their products, monetisation models and marketing strategies. 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.

Nonetheless, the interviewed app developers noted that there are important limitations in what insights they can reach without help and data sharing by the app stores. Several app developers explained that they need to understand how end-users engage with similar apps, and what makes the users like them. For example, developers of video games need to understand how to make a game neither too easy, not too hard for the players. Data generated through their app alone cannot provide a comprehensive answer to this question, and therefore a broader overview of similar apps and user behaviour in that segment is desired.

In addition to this, similarly to businesses in other sectors, app developers make use of consumer behaviour data, business analytics, market trends analysis, and so on. This concerns analysis of their own business performance, as well as competition.

---


3.2.3. Accommodation and hospitality industry

Based on the interview and desk research findings, the business users of hospitality sector have relatively lower data needs, as compared to e-commerce sellers and especially app developers. 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. Hotels compete for customers within a limited geographic area, 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. However, considering the data needs, the distinction should be made between smaller properties and large hotel chains.

The key information that large, multi-city hotel chains are interested in is the customer contact data. It is the hotels’ chance to build guest relationships with loyalty marketing. It is also crucial in targeting the right profiles with promotions and rewarding the best guests with campaigns that encourage loyalty - 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. These facilitate for 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. Smaller properties, on the contrary, do not have a goal to bypass OTAs, as they allow the properties to be visible for broad audiences of travellers from around the world. Therefore, the need for customer contact details is not as pronounced in this size segment.

Besides contact data, properties of various sizes work with several areas of indicators. As a starting point, knowing the competition is key to attract customers and revenue management. 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. In addition to this, analysis of market trends (e.g., seasonal trends, trends related to external events, such as 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 and, more importantly for large hotels, to drive engagement, guest loyalty, direct booking, and ancillary revenues.

In addition to this, the interviewed hotel owners mentioned that they need data to deal with fake or undesirable reservations. The platforms are reluctant to provide such data, because it usually concerns personal details of customers, such as the IP address, email validity, or reservation success on the platform.

Large hotel chains, as compared to small hotel owners, often have own capacities to source and work with large amounts of raw data, draw insights, and base large-scale business decisions on those insights. These include customer preferences, marketing and pricing strategies, inventory management. The

---

67 For example, Hilton hotels have a detailed framework for analysing negative reviews and drawing managerial implications, see https://www.slideshare.net/georgeskiadopoulos5/analysis-of-negative-hotel-reviews-hilton-hotel
analysed data usually come from other sources than the online platforms that they use to list the properties. This is related to the fact that usually limited shares of their rooms are listed on online platforms.

3.3. General data needs of platform business users

Several broader insights on business user data needs could be made based on the evidence collected in preparation of this analytical paper. Rather specific types or characteristics of platform data (please see the typologies presented in Section 3.1) are relevant for businesses across all reviewed sectors.

First, all interviewees mentioned that they mostly use and appreciate the data on the individual business user’s performance, listings, advertising effectiveness and so on. This is essential in managing the online business. Nonetheless, data on user identification, transactions and performance of other businesses is also sought after by business users operating in various sectors and on various platforms.

Second, based on the insights from interviews and desk research, the need for data on competition and market trends seems to mostly concern certain types of businesses across sectors, specifically those looking for opportunities to scale up.

Third, most business users indicate that they need processed data - analytics and insights, rather than raw datasets. Most interviewees acknowledged, that they would not know how to analyse data for business insights themselves. Meanwhile, almost a third enterprises that responded to the business user survey do not analyse any kind of raw data (see more details in Section 5.2). Nonetheless, many players in the data market can offer such analytics services for companies operating in various sectors - we elaborate on this issue in Chapter 6.

In general, the discussion about data access, especially among retailers and hotels is still rather vague and sometimes contradictory. Many businesses do not know what data is (or can be) collected by platforms, and how could they use this data to get competitive advantage. There is no shared vision among business users in different sectors on how the data could be shared and under what circumstances. This could be attributed to both, lack of data savviness of businesses, as well as the general lack of knowledge on what data platforms collect and what specific metrics they monitor.

The evidence on platform business user data needs, collected in preparation of this paper, also indicated some unrealistic expectations of business users. To begin with, there is lack of understanding among the business users what data can be legally shared. This especially concerns the notable tension between the business users’ desire to have access to more customer and competitor data on one hand, and the GDPR and anti-trust law on the other. At the same time, open access to some data could be detrimental to the consumers (e.g., flooding with advertisement) or competition (e.g., competitor sales data with identification information) more generally. Further, not all data is feasible to be provided by the platforms due to other barriers. We elaborate on the platforms’ perspective and barriers to data sharing more in Chapter 4 and Section 4.5 specifically.

In the following chapter, we analyse in more detail the specific data collection and sharing practices of the four selected online platforms: Amazon, eBay, Google Play and Booking.com, as well as provide a broader overview of data sharing practices in the platform economy.
**4. Platform data and business user access**

Online platforms are in the best position to collect data and capture its value, given their capacities to invest in infrastructure, notable data advantage against other parties in the platform economy, and lack of knowledge among internet users on how to value their own data. The users (both customers and businesses) provide data actively and passively. For example, browsing the internet they generate a wealth of search and user logs, customer transaction records, user-generated content, and rich network information. In addition to this, major platforms, such as Amazon and Google have developed a range of gadgets (e.g., Amazon Echo, Google Nest) and operational systems linking phones, personal computers and smart home appliances (e.g., Alexa, Google Home, Apple HomeKit), allowing to collect the data on customers outside the ‘traditional’ web-browsing environment. Internet of Things, 5G networks, and the emerging online-to-offline transition are further accelerating the speed of data accumulation.

The basic model of how platforms collect and derive value from data is similar across all sectors of the platform economy (see the figure below): the data is generated by the activities of end-users (customers) and third-party business users on the platform. This data is then analysed by the online platform companies to develop data-based products or services, or used in making their business decisions.

**FIGURE 7. DATA SOURCES AND CREATION OF THE VALUE OF DATA**

Nonetheless, what data exactly the platforms collect, how the data flow within online platform networks, how they are monetised, and how their value is captured, can differ significantly by the platform business

---


68 Google Store. Retrieved from [https://store.google.com/ie/?hl=en-IE&regionRedirect=true](https://store.google.com/ie/?hl=en-IE&regionRedirect=true)

model (e.g. degree of vertical integration in the data value chain)\textsuperscript{70}. Differences can also be observed in the practices of sharing their data with business users, who are key in generating them.

Further in this section we provide a detailed overview of the data that the selected platforms - Amazon, eBay, Google Play and Booking.com - collect, and what level of access to this data they provide to the business users. Then we present a broader overview of platform motives to provide or restrict access to their data.

4.1. Amazon Marketplace

Media articles often present Amazon as the largest retail data repository in the world\textsuperscript{71}. Amazon collects data from customers as they navigate the site or use other Amazon’s products or services. As the platform itself informs, they monitor the following:

- Actions such as searching for products or services, placing an order, submitting a review or compiling a Wishlist gives Amazon all details on browsing the site, from the time customers spend browsing each page to the language used in the user reviews they leave.
- When customers are downloading, streaming, viewing, or using digital content on a device, or through an app, especially on a GPS-enabled smartphone or tablet, Amazon can also gather location data and information about other apps and services they use on the phone. Using Amazon’s streaming content services, such as Amazon Prime and Audible, provides them with more detailed information on where, when and how you watch and listen to TV, film and audio.
- By providing or configuring settings, Account or Profile information, customers reveal their name, address and phone number, age, personal description and photograph, signature, payment information, privacy preferences.
- Talking to Amazon’s Alexa virtual assistant allow to gather customer voice recordings that are stored by the platform.
- Communicating with Amazon by phone, e-mail, or otherwise provides data on the content of reviews, calls and e-mails\textsuperscript{72}.

Some evidence also shows that the platform invests in expensive experiments for data collection. For example, according to some reports, to gather information of consumer price sensitivity, Amazon runs experiments that involve funding discounts on third-party products, and exploring customer behaviour in relation to decreased prices\textsuperscript{73}. While occasional anecdotal evidence is provided by various whistle-blowers (often the platform’s former employees), the full extent of the collection of customer data by Amazon and the methodologies applied are not disclosed publicly.

The platform has a lot of information on the businesses registered on the platform as sellers as well. Besides all the activities and transactions on the platform, Amazon collects product information, insurance data, trade licenses (for service providers), contacts, bank and tax information, as well as


\textsuperscript{72} Amazon. (n.d.). What data does Amazon collet and use? Retrieved from https://www.amazon.co.uk/gp/help/customer/display.html?nodeId=G6RZ4RMNMLUQRLY2

\textsuperscript{73} Financial times. (n.d.). Amazon’s ever-increasing power unnerves vendors. Retrieved from https://www.ft.com/content/cf2ce968-bc8a-11e8-94b2-17176eb93f
business demographics, such as sector, business type, registration address, year of business and number of employees. Linked with the customer data, this results in a wealth of information on how the market works.

The possibilities to analyse this data to draw commercially-viable insights are basically unlimited. Interviews with business users and desk research indicate that the platform has information ranging from the click-through rates of each listing, paths through which customers end-up viewing specific products, complementarity of products and price level adjustments24 (none of which is shared with the sellers), to very detailed information on individual customers and their shopping habits. The platform also makes use of external datasets, such as census data for gathering demographic details75, which can help to develop a very detailed profiling of each customer. Through advanced analytics, Amazon has a clear understanding of what people actually buy76, search for and do not buy, and what they may buy next - this gives it an extreme competitive advantage37.

The platform shares only a very limited amount of this knowledge with its third-party sellers, especially those who do not use Amazon’s additional services, such as fulfilment and advertising. As the interviewed Amazon sellers reported, this becomes even more of an issue because of the platform’s vertical integration: Amazon Retail - often their direct competitor on Amazon Marketplace - knows everything about its competition and its competitors.

To give a little bit more context, Amazon Marketplace P2B relationships can be described by the distinction between vendors and sellers. Vendors are retailers that sell wholesale to Amazon Retail (first party), while sellers are third parties operating on Amazon Marketplace. Vendor relationship with Amazon is somewhat similar to that between a supermarket and its suppliers. Generally, vendors are outside our definition of platform business users, but their example offers interesting insights into Amazon’s data services for businesses.

Businesses that become direct vendors to Amazon, with Amazon Retail Analytics (ARA) Basic, get elementary data, such as sales and inventory data, weekly. However, this package does not include more in-depth data that vendors would like to have, such as conversion rates or customer details. The 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 contact78. Brands that sell through Vendor Central have to pay for a subscription to Amazon Retail Analytics (ARA) Premium to access more data at a higher level of granularity. However, it is expensive, with vendors reportedly paying 1% of their wholesale cost of goods sold to Amazon or a minimum of USD 100,00079 (although some other sources also indicate smaller prices, such as $30,000 per year80) to get access to a database that lets them see some, but not all, of the data Amazon has compiled81. Also, some reports from former Amazon

---

27 Financial times. (n.d.). Amazon’s ever-increasing power unnerves vendors. Retrieved from https://www.ft.com/content/c82ce968-bc8a-11e8-94b2-1717ebfb935
employees show that some larger vendors may be provided with more information on case-by-case basis, if they get a green light from form Amazon superiors. Interestingly, an interviewee representing several large businesses active on Amazon mentioned that sometimes vendors on Amazon also become third-party sellers in order to get an access to more data.

**FIGURE 8. AMAZON RETAIL ANALYTICS: BASIC VS PREMIUM PACKAGE**

For third-party sellers, especially those who own a brand, Amazon offers more data. In contrast to the Amazon vendor model (or similarly to a brick-and-mortar supermarket), when a consumer purchases goods online through the Amazon Marketplace, not only Amazon but also the third-party seller can acquire some consumer data. Nonetheless, 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.

Further, sellers on Amazon get access to some conversion metrics — essential for improving content on their product pages. They can retrieve (including through an API) information on their own business performance, 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, as presented in Figure 6) for the following categories:
  o Advertising: Recommendations for advertising your products and for using the sponsored products program.
  o Fulfilment: Recommendations for the top products to fulfil through Fulfilment by Amazon (FBA).
  o Global Selling: Recommendations for expanding your products to more regions and marketplaces.
  o Inventory: Recommendations for restocking low or out-of-stock items in your inventory.
  o Pricing: Recommendations to review pricing on items in your inventory where your offer is not the lowest price.\(^{86}\)

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

### TABLE 1. AMAZON BRAND ANALYTICS, AS ACCESSED BY A SELLER

<table>
<thead>
<tr>
<th>Amazon Search Terms</th>
<th>Amazon Search Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>• For each search term:</td>
<td>• For each search term:</td>
</tr>
<tr>
<td>o department (.de, .co.uk, .com, etc);</td>
<td>o department (.de, .co.uk, .com, etc);</td>
</tr>
<tr>
<td>o search frequency rank;</td>
<td>o search frequency rank;</td>
</tr>
<tr>
<td>o #1, #2 and #3 most clicked product (by Amazon Standard Identification Number - ASIN) for that search term, with its click share (%) and conversion rate (%).</td>
<td>o #1, #2 and #3 most clicked product (by Amazon Standard Identification Number - ASIN) for that search term, with its click share (%) and conversion rate (%).</td>
</tr>
<tr>
<td>o No actual search volume or volume of purchases in absolute terms.</td>
<td>o No actual search volume or volume of purchases in absolute terms.</td>
</tr>
<tr>
<td>• They can be filtered by reporting range (e.g., weekly, specific week).</td>
<td>• They can be filtered by reporting range (e.g., weekly, specific week).</td>
</tr>
<tr>
<td>• Could be downloaded in excel or CSV or explored in a web browser.</td>
<td>• Could be downloaded in excel or CSV or explored in a web browser.</td>
</tr>
</tbody>
</table>

86 Amazon Services Europe. What you should know about the Amazon MWS Reports API section. Retrieved from http://docs.developer.amazonservices.com/en_UK/reports/Reports_Overview.html
• *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).
  - 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 volumes of products sold*

**Demographics**
- Breakdown of the seller’s Amazon customers (in aggregate) by age, household income, education, gender, and marital status, by a purchased product brand or category, but not by a specific product.

In addition to the Brand Analytics Tool, businesses using Amazon’s paid advertising tool can get additional data and analytics, such as ACoS (advertising cost of sales), showing the ratio between advertising costs and sales revenues. Moreover, in autumn 2019, the platform released a beta for advertising attribution for marketplace sellers. Amazon Attribution is a measurement tool that tracks the performance of Amazon ads on channels off Amazon. The attribution measures page views, purchases, and sales driven by external campaigns.

The table below broadly summarises seller access to the categories of data and analytics provided by Amazon, using the typologies presented in Chapter 3. Generally, according to most interviewed Amazon sellers and desk research, information on an individual seller’s operations and performance that the platform provides is useful and sufficient for simple insights. Interviews with Amazon sellers and desk research also indicated that generally the seller community appreciates the Brand Analytics Tool, as it is free, allows, at least to some extent, to “keep an eye” on the competitors, helps to apply potential keywords, and provides better knowledge of the target audience and products. As we discuss in more detail in Chapter 6, this data becomes especially valuable when integrated with sales data - a feature provided by third-party analytics companies (e.g., Jungle Scout).

**TABLE 2. TYPES OF DATA AND ANALYTICS ACCESSIBLE BY AMAZON MARKETPLACE SELLERS**

<table>
<thead>
<tr>
<th>AMAZON 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>Own</td>
<td>Own</td>
<td>Own</td>
<td>Own</td>
</tr>
<tr>
<td>Insights/Actionable guidelines</td>
<td>Raw data</td>
<td>Analytics</td>
<td>Insights/Actionable guidelines</td>
<td>Raw data</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>Some</td>
<td>NA</td>
<td>Some</td>
</tr>
</tbody>
</table>


Nonetheless, some interviewed sellers argued that the extent, comprehensiveness and granularity of data provided is still very limited. Brand Analytics, according to them, is a miniscule part of the full picture. Most of the interviewed sellers indicated that they, as well as other sellers they know, widely use third-party data and analytics tools to compensate for the lack of information on the marketplace trends and activities. Some interviewees also argued that Brand Analytics is by far not enough, especially considering what access to data Amazon Retail has.

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”\(^9\). 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)\(^9\). Recent media articles provide a wealth of stories on unfair competition between Amazon private labels and third-party Marketplace sellers\(^2\), using the aggregated seller data. Amazon’s practices of sharing seller data between Retail and Marketplace, as well as sophisticated use of metrics to boost its own sales was also corroborated by the platform’s former employees\(^3\). In November 2019, Amazon itself confirmed that it uses “aggregated data” from sellers in its third-party marketplace to improve its overall business\(^4\).

Indeed, as pointed out by the EU and US policymakers\(^5\), Amazon’s data advantage over its business users constitutes a key problem with policy relevance. As a platform in a dominant position (in terms of its reach; around 200 million customers each month\(9\)), Amazon can use the data to identify successful products with a corresponding price, then offer them itself, at a better price or delivery conditions.

Meanwhile, the platform itself argues that third-party sellers generate an important and increasing share of its revenues (third-party sales amounted $160 billion globally in 2018\(^9\)), so it has no incentive to harm the sellers - on the contrary, it invests in tools to help their businesses scale. Amazon’s private label products, according to an interviewed Amazon representative, 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\(^8\), others estimate that the private label business totalled in $7.5 billion in 2019\(^9\) 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\(^10\). However, some sources estimate that Amazon’s

---

\(^9\) Dayton, E. (2020). “Amazon statistics you should know: opportunities to make the most of America’s top online marketplace”. Available at https://www.bigcommerce.com/blog/amazon-statistics/?a=shopping-experience-beyond-compare
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)\(^{103}\). Meanwhile, an American investment bank projects that Amazon’s private label business could see revenues of $25 billion by 2022\(^{102}\). The data that Amazon collects on all the transaction on the Marketplace contributes significantly to this.

### 4.2. eBay

Compared to Amazon, eBay, active in the e-commerce segment of platform economy, has fewer users, as well as a simpler business model (without vertical integration) in which the core source of its revenues is the marketplace. As of 2018, eBay was significantly smaller than Amazon in terms of global third-party sales ($95 billion)\(^{103}\).

The total scope of data collection, therefore, is significantly smaller than Amazon’s 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. Some of this data is actively provided by a user when using the platform or creating an account (e.g. identification data, financial information, contacts); while some of it is collected automatically when a user uses the platform or creates an account (e.g. transactions, location data, computer connection data). The platform also uses cookies to track customer activities (e.g. page visits, user segment, device information). 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);
- credit agencies or bureaus (e.g. credit reports/checks);
- data providers (e.g. demographic, interest-based and online advertising related data).\(^{104}\)

In data analysis, eBay deals with structured, unstructured, and semi-structured data, where Hadoop, as a big data platform, provides key technology features\(^{105}\). The company is also said to be actively contributing its knowledge back to the open source community\(^{106}\).

Similar to Amazon, eBay provides a number of analytics metrics to its business users though eBay Seller Hub. Similar to Amazon, the platforms’ analytics package is sometimes criticised for not being flexible, granular or extensive enough\(^{107}\). While in 2018 the range of seller tools provided by eBay - including data

---


---

34
and analytics - was slightly broader than provided by Amazon,\textsuperscript{108} this trend changed with Amazon introducing its Brand Analytics in 2019.

The most important metrics that sellers can subscribe to on eBay (Listing Analytics, on individual listings, as well as overall averages) for free\textsuperscript{109} include:

- Sales history, changes in sales, overall performance over time.
- Selling costs, including specific eBay fees, shipping labels, PayPal fees.
- Competitor’s pricing (also in comparison to the seller’s pricing) and some competitor performance data, in also comparison to seller’s performance.
- Listing data, including ranks, click-through rates, page views, sales conversion rates, transactions, where they appear in search results, how many people see them in results\textsuperscript{110}.

eBay business users can access this information through their profile in the eBay Seller Hub or an API connection. In fact, eBay seller community claims that Listing Analytics is the most popular eBay Apps Centre application of all time. Sellers report it to be free, simple to use and generally useful.

In the recent years, eBay has bought and incorporated third-party analytics applications such as Terapeak\textsuperscript{111}. When the acquisition 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 to attract their business. 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\textsuperscript{112}.

Terapeak currently powers the Listing Analytics tool, as described above. Moreover, for sellers with a Basic, Premium, Anchor or Enterprise Store, an additional product Terapeak Research is included in the subscription free of charge. All other sellers using eBay (basically Starter accounts) have the option to pay for a yearly or monthly subscription\textsuperscript{113}. We discuss this and similar third-party data and analytics products in more detail in Chapter 6. The table below broadly summarises seller access to the categories of data and analytics provided by eBay, using the typologies presented in Chapter 3.

### 4.3. Google Play app store

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. 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 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 are conducted on the issue\(^\text{114}\), and various whistle blowers occasionally bring to attention new ways and areas in which the platform collects user data\(^\text{115}\). Some findings indicate that Google has the ability to connect the anonymous data collected through passive means with the personal information of the user\(^\text{116}\).

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\(^\text{117}\), it is difficult to strictly delineate data that is collected exclusively on Google Play from data generated in Google’s other activities. 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 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.

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) as well as extensive material on how to use this data in decision-making. The main categories are detailed in the box below. Some of the data on apps of Google Play is also accessible publicly, and is being used in academic research. Some developers report that this information enabled them to improve customer retention, engagement and monetisation.

TABLE 4. 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. This data allows you to judge the performance against key areas of the ecosystem.

---


Monetization: Google Play also provides 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 have a very good understanding how the developer is doing. 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, 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.²²⁰

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 of fraud detection. Although Google Play is a vertically integrated platform (i.e. it distributes Google 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.

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 it is up to developers themselves. 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 policy²²¹, 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\(^\text{122}\). Google said they would address the issues in Android 10, which was released in September 2019\(^\text{123}\).

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

**TABLE 5. 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 Competitor</td>
<td>Own Competitor</td>
<td>Own Competitor</td>
<td>NA</td>
</tr>
<tr>
<td>Raw data</td>
<td>Depends on user permissions</td>
<td>Yes No</td>
<td>Yes No</td>
<td>No No</td>
<td>NA</td>
</tr>
<tr>
<td>Analytics</td>
<td>NA</td>
<td>Yes No</td>
<td>Yes Some</td>
<td>Yes Some</td>
<td>Yes</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>

4.4. Booking.com

Booking.com, a hotel booking platform, is one of the major European-based online platforms in the world. Much like all the 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 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”\(^\text{124}\). 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\(^\text{125}\).

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\(^\text{126}\) using the Booking.com platform. This, in turn, also provides the platform with insights on destinations and countries all around the world.

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. 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\(^\text{127}\).

Furthermore, Booking.com conducts analytics on these data to provide third-party sellers data targeting 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\(^\text{128}\). According to the platform, the aim is to help hotels boost your sales and find ways of getting more bookings. The box below provides a detail list of items that Booking.com provides to its business users in the Analytics section.

**TABLE 6. 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 information on the hotel’s current bookers, helping to better understand who their customers are\(^\text{129}\);  
| • 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\(^\text{130}\) (~10-15 properties).
• “Genius” programme reports (for participating properties), including statistics on Genius guests and gains (additional revenues) from the programme\(^\text{131}\).
• Visibility Dashboard, providing data to track and improve property’s visibility in Booking.com search results, such as:
  o search result views (the number of times potential guests saw the property in search results);
  o 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);
  o property page views (the number of times potential guests viewed the property page);
  o bookings (the total number of bookings received, including cancellations) and the overall conversion rates.
  o visibility performance in Booking.com search results the past 30, 90 or 365 days.
  o comparative visibility performance against other properties in the market for the past 90 days (see the figure below).
  o Conversion (how well the property converts guest searches into bookings).
  o 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)\(^\text{132}\).

---

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\textsuperscript{33}, providing market data (demand, room rates, local events) for informed pricing decisions\textsuperscript{34}. It was reported that Booking.com data analytics service on pricing strategy on average increased third-party sellers’ sales revenue by 7\%\textsuperscript{35}. 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 (e.g., impact of Brexit on hotel bookings).

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 communication tools\textsuperscript{36}. With an OTA booking, hotels are not passed along any first-party data, leaving the only chance to capture customer contact data at the point of sale\textsuperscript{37}. It is difficult for hotels to capture this information at the front desk since it is a request rather than a requirement (as it is for registering and making reservations on Booking.com). The customer contact data, meanwhile, is hotels’ chance to build guest relationships with loyalty marketing and promotions\textsuperscript{38}.

Therefore, while the interviewed hotels reported that they appreciate the analytics and insights that Booking.com generates for them, they lack 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, “hosts are aware in advance and can serve customer better”. Nonetheless, smaller establishments on Booking.com, as reported by Booking.com and other interviewed platforms in tourism sector, are less likely to use these insights at all.

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 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>Data on transactions between businesses and customers</td>
<td>Own</td>
<td>Competitor</td>
<td>Own</td>
<td>Competitor</td>
</tr>
<tr>
<td>Raw data</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>
</tr>
<tr>
<td>Insights</td>
<td>NA</td>
<td>Some</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

\textsuperscript{133} While larger properties with own IT departments can get this data through an API and analyse it themselves.


\textsuperscript{137} 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.

4.5. Reasons for providing data or restricting the access

As the four platform case studies show, the kinds of data provided and not provided by these players are rather similar. First, all the overviewed platforms have made notable investments in analytics capacities to draw insights from this data and provide it to the business users. Therefore, all business users get access to some data and analytics important to their commercial activities. However, they do not share the raw big data, as well as detailed data on customers and competitors. Although platforms reportedly collect and analyse loads of data relevant for competition on their marketplaces (basically everything that takes place on their apps and websites every second), only a fraction of this usually is shared.

Online platforms provide several main justifications for not granting the access to data (or a combination of them), that could be summarised into several main categories:

- **Protection of business model.** For example, OTAs argue that if they provide customer contact data to establishments that are listed on these platforms, the establishments will be able bypass them, destroying the source of revenue. 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.

- **Protection of trade secrets.** As an interviewed marketplace representative of a European platform explained, data openness is risky. Often, they detect abusive bots originating from China, crawling their pages or trying to use their APIs. At the same time, they see Chinese marketplaces as serious competitors, disregarding all the regulations they must comply with, and therefore gaining competitive advantages.

- It is useful to analytically distinguish between data as basis of the business model and data as by-product of activities on the platform. Although the latter category of data could be easier for platforms to share, they point out other challenges, such as data interoperability. In many cases, the obstacles to interoperability are technical. Overcoming this to develop new interoperability functionalities, is costly and technically complicated according to a representative of a developer’s organisation. However, often interoperability issues arise due to misaligned incentives as such interoperability might have broader benefits but to the cost of the dominant companies. Some platform companies, nonetheless, are already making substantial efforts in this regard, such as the Data Transfer Project that includes Microsoft, Google, Facebook and Twitter.

- The provision of raw data generated by the platform is problematic, both because of its volume and usability. To be useful for business users, most types of data have to be analysed and be used to provide insights. Meanwhile, business analytics and insights, which could really benefit the business users, require significant investments from the platforms. From platform business point of view, such investment could not pay off.

- **Lack of need to provide data.** Related to the previous point, platforms often argue that the majority of business users would not know what to do with data and how to make it useful. In fact, many interviewed business users who reported feeling disadvantaged by limited data access, could not provide a detailed vision of how they would analyse the raw data from

---


140 Data Transfer Project. (n.d.). Retrieved from https://datatransferproject.dev/
platforms and draw insights. On the other hand, the interviewed large companies using online platforms claim that this justification is a “nonsensical excuse”.

- **Company status.** Platform companies that are privately owned (e.g., Allegro) are not required by law to disclose detailed financial and operating information in most instances. They enjoy wide latitude in deciding what types of information to make available to the public and have the flexibility to release their data whenever they need, e.g. real-time. Meanwhile, US publicly traded companies (e.g. Amazon) are obliged to make market-sensitive information available to all parties at the same time\(^\text{141}\). This allows Allegro to share their real-time gross merchandise volume (GMV) data at any time, while Amazon - a publicly traded company - cannot do this because of the disclosure laws and requirements that apply to it.

- **Data protection legislation.** The interviewed platforms seemed to be rather concerned with meeting the requirements of GDPR, while some of the business users see this as a justification for not providing customer data to them. In fact, from the legal point of view, this concerns personal data only to the extent that they cannot be anonymized. Moreover, the platforms could comply with the requirements of GDPR by giving their business users the opportunity to obtain their own consent from individual customers. Besides personal data, platforms also must protect the platform business user data in order to comply with competition policy rules. For example, a representative of an e-commerce platform explained that sharing information on the revenue of a specific seller active on the platform with other sellers can be interpreted as providing a competitive advice under the German anti-trust law.

- **Customer satisfaction.** All the interviewed platforms emphasised customer satisfaction as their key goal of their policies. E-commerce and accommodation platforms expressed a concern that if the sellers get customer email information, they will flood the customers with marketing messages. Meanwhile additional requests for data might confuse the customers (e.g., at the moment of hotel check-in or during the transaction), and this is how platforms justify not providing business users the possibility to obtain customer consent to process their data themselves.

From the economic point of view, the incentives for companies to grant access to their data mostly depends on the role of data in their business model. If the exploitation of a particular dataset that it collects is at the core of its business model, an online platform will be unlikely to share it\(^\text{142}\). Data and its analysis form the core of online platforms business models, which could explain the general reluctance of these companies to share their data\(^\text{143}\). As explained by an association representing several platform operators, data is a very sensitive question for them.

Nonetheless, in some cases sharing of data is beneficial. Data access for business users, as well as expensive advanced analytics and tips, are granted to the extent that it could be generate more income for the platform. First, it is in the interest of platforms to allow the business users develop better products and services, attract more customers, conduct more transactions and therefore generate more income for the platform through commission on intermediation services. Second, platform business users also often want more data, so if a competing platform invests in analytics, others are motivated to do so too to maintain their business user base. Most interviewed platforms emphasised these as the key incentives to


invest in analytics capacities for business-user oriented analytics and insights. One of the clearest limits to data provision is any kind of data that would allow businesses to reduce dependency on the intermediately platform by shifting the traffic to personal online stores or bypassing the platforms overall (i.e., customer contact/identification data). Meanwhile, vertically integrated platforms want to avoid providing data that would allow business users to become serious competitors. Depending on other variables, such as investment and competition, these “limits” of data provision could be expanded or lowered.

It is important to note that the four analysed cases do not fully reflect the full variety of data sharing practices in the platform economy. This overview did not allow to identify notable data sharing differences by sector or between vertically integrated and non-integrated platforms — a larger sample comparative study is needed to determine this. For example, the vertically integrated German platform Real.de does not provide its sellers with any kind of comparative analytics or data on the overview of other user activities on the platform, unless they’re using platform’s advertising services. Meanwhile, the Polish marketplace Allegro engaged purely in 3p activities (i.e., non-integrated) provides all the data on transactions real-time through its API, including individual-level anonymised data on all the sellers. However, these differences may as well be determined by other factors than vertical integration, related to specific contexts, a platform’s mission and vision, as well as decisions made by platform companies.

---

144 To determine this, more in-depth analysis into the functionalities provided by the platforms to their business users is needed. The desk research generated some but not sufficient evidence. For example, eBay shares more data with third party data and analytics provider than Amazon; eBay and Etsy allow to connect the business account to Google Analytics, while Amazon does not. This information can be retrieved from https://community.ebay.com/55/Archive-Selling/How-To-Use-Google-Analytics-To-See-your-Store-Traffic-Stats/id-p/24230195; https://help.etsy.com/hc/en-us/articles/36000337967-Setting-up-Google-Analytics?segment=selling; and https://sellercentral.amazon.com/forums/t/can-i-use-my-google-analytics-account-to-track-stats-on-my-amazon-shop/212291.
5. Business user perspective on data access

In this chapter, we briefly present the business user perspective on the data that they receive from the platforms, alternative data sources and strategies to acquire data. We build on data from desk research interviews and the business user survey to present a general overview of their views towards platform data access. Ultimately, this helps to better understand the key gaps between the business user data needs presented in Chapter 3, and data and analytics provided by platforms, presented in Chapter 4. In the second part of the chapter, we discuss the options that business users have to fill these gaps.

It must be noted that the business user survey, results of which are presented in this chapter, targeted the business users of all internet platforms that fit the study definition, not only e-commerce sellers, accommodation providers and app developers. More specifically, the enterprises that participated in the survey use platforms for one or several of the following purposes: selling goods, selling services, distributing apps or software, listing accommodation and advertising. Therefore, the survey results reflect a broader overview of data sharing in P2B relationships.

5.1. Business users’ perspective on platform data access

A notable share of business users of online platforms report that they do not get satisfactory access to data from platforms - both in the earlier studies, as well as the survey and interview programme conducted for this paper. For instance, Ecorys study showed that 33% of the ‘heavy’ users of online platforms responding to a survey stated that they had experienced data access related issues in their trading practices with online platforms. Meanwhile, our business user survey results show that around 27% of surveyed enterprises disagreed, strongly disagreed or partly disagreed with the statement that their business can easily access data collected by the main platform, which is important to their business (Figure 9).

145 According to a study by Ecorys, 26% of surveyed business users said they had access to data on transactions and market analysis. Only 21% said they could access user behaviour data. Access to the data of other users, e.g. potential customers, is granted only in rare cases. Furthermore, according to a Eurobarometer study (Eurobarometer 439), 42% of surveyed SMEs reported not getting the data they needed from online marketplaces.

Further, quite in line with the business user interview and platform analysis data presented in Chapter 4, the majority of business users report to receive data on their performance and their customers (see the figure below). Meanwhile, significantly smaller shares of businesses report also having access to data on other businesses and other customers.

Most of the respondents receive the platform data in aggregated from or summarised and visualised into reports - without any data at individual level (see the figure below). Meanwhile, over a third of respondents also receive raw, individual level data. Interview results allow to interpret this as primarily the data related to individual transactions of the business user, which is provided by all platforms.
FIGURE 11. HOW MUCH PROCESSING DOES THE MAIN PLATFORM CARRY OUT ON DATA IT PROVIDES TO YOUR BUSINESS?

![Diagram of processing levels](image)

Source: PPMI; n=1,500

Generally, platform business users (participating in interviews and voicing their positions on the internet) claim that they would like to have more of both: individual level data about their own operations or own customers/consumers, as well as aggregated data on the entire market including, information on competing business users’ operations and their customers/consumers. Also, in a more detailed survey question, around a third of the respondents indicated that they cannot access more specific kinds of data held by the main online platform that they use, which are essential to their business (Figure 12). This concerns indicators under most data types, as the variation across the different data types is marginal.

FIGURE 12. NEEDED BUT INACCESSIBLE DATA HELD BY THE ONLINE PLATFORM

![Bar chart showing data needs](image)

Source: PPMI; N=1,667

Among those survey respondents, who indicated that they cannot easily access the data (Figure 9), the most frequently perceived difficulty in accessing data from the main platform is legislative or regulatory restrictions (Figure 13), followed by prohibitively high prices for accessing the data.

---

147 Those who disagreed, party disagreed or strongly disagreed.
In fact, in another survey question, around half of respondents reported that their main platform charges additional fees for access to specific kinds of data. Around 60% of these businesses (N=877) reported paying up to €5,000 for data annually, while around 40% - more than €5,000.

Finally, among the surveyed businesses who have experienced problems, disputes or disagreements with their main platform (N=769), 16.5% reported lack of access to data as a key reason. This indicates that in some cases restricted platform data access can be an especially important issue for businesses.

5.2. Data alternatives for business users

Platform business users have a range of options to collect data and generate business intelligence outside the online platforms, as well as triangulate data sources to build more valuable insights. This depends on several aspects.

First, that is analytical capacity. As explained in Chapter 3, different business users have very different levels of data savviness and ability to collect and draw actionable insights from raw data. For example, capacities of app developers can be hardly comparable to that of small hotel owners; larger businesses in all analysed sectors tend to invest more in own IT departments and data analytics to develop analytics and business insights. According to results of the business user survey, the majority of the companies reported that they analyse some kinds of raw data: over 42% of enterprises do it in-house, 25% outsource such analysis to external service providers (Figure 14). Around 7% of companies both analyse raw data in-house and outsource. Meanwhile, almost a third of surveyed enterprises do not analyse any kind of raw data.

---

**FIGURE 13. DIFFICULTIES EXPERIENCED IN ACCESSING DATA FROM THE PLATFORM**

![Bar chart showing difficulties in accessing data.](chart)

Source: PPMI; n=438.

148 Those who responded in survey Q10 that they cannot easily access data (i.e. disagreed, strongly disagreed or party disagreed with the statement ‘My business can easily access data collected by the platform that is important for my business’). See questionnaire in Annex 2.

149 Please note that after a closer inspection of the survey results, the research team considers that some of the respondents had in mind not only the fees charged by platforms, but also the fees of third party providers, offering insights on the specific platforms.
The second key factor is financial capacity. Emerging models of synthetic data for training Artificial Intelligence applications and cloud solutions offer new, innovative approaches to data challenges, avoiding the need for accessing the vast volumes of data collected by platforms, and sometimes competing with the platforms’ own applications. Certainly, access to such innovations comes at a cost, so analytical capacities often depend directly on the availability of financial resources.

The third key precondition for accessing the needed information for businesses is the existence of specific data outside the platform. While acknowledging the data asymmetry between platforms and their business users, some competition researchers argue that data markets are active and thriving. Most of these data are actively traded and accessible to anybody willing to pay for it. Nonetheless, opponents of this view claim that specific data is not necessarily available on the market, and most of the data that platforms collect and not share is unique to specific activities going on those platforms and cannot be duplicated. Such non-duplicable data include:

- Total data on overall commercial transactions on the platform: the total database of transactions on the platform cannot be collected by individual sellers or other players (platforms, data brokers) in the platform ecosystem.
- Data on real-time of activities on the platform (e.g. X currently has product Y added to the shopping cart), which allows to effectively address the customer.
- Transaction-related data about the customers, sales activities and listings of specific business user also cannot be duplicated by that user, as the data is recorded exclusively by the platform on which the transaction takes place.
- Analytics developed by a platform and based on the data that it has also cannot be duplicated at the same level of precision by other players.

As an interviewed Amazon seller explained, they need data on the specific platforms that they use. Increasingly, the market power is concentrated within one or two large platforms, so the sellers do not have much choice on which one to sell if they want a successful business (“Amazon has a 60% market share in Germany, you cannot just go sell on eBay and be OK; eBay is marginal in Europe”). If the major platform refuses to share its data, this is something that its business users must deal with.

Nonetheless, business users have two broad ways to get access to additional data or analytics: (1) to manually collect and analyse data from sources available to them; (2) rely on third party providers.

---

150 Several start-ups creating, for their own use, or selling synthetically generated data for training AI models are emerging, some of which are successful European-based companies. See, for example https://www.wired.com/story/some-startups-use-fake-data-to-train-ai/


153 Interview with an association representing large businesses that use online marketplaces.
Although innovative approaches to acquire data are emerging (see the box below), most platform users apply one of these or a combination of them.

TABLE 8. COOPERATIVE PLATFORM AS A WAY TO GENERATE DATA

A report by the German Economic Institute\(^\text{154}\) suggests that the model of a cooperative could be particularly well-suited to organise sharing of data among e-commerce traders. Data knowledge-based services could be made available to members of cooperatives who, for example, build a digital platform that members can use to offer their goods and services. One such example is Fairmondo.de\(^\text{155}\), a platform which belongs to its business users and employees. Through a cooperative structure, the users can share the platform as a resource for mutual benefit without being dependent on large e-marketplaces.

Authors of the report suggest that the data accumulated on a cooperative platform, and thus also the economic benefits derived from these data, would be available to all members who contribute to the data creation process on the platform, rather than to the central platform provider only. Meanwhile, the practical usability of data, which is often difficult for individual providers to achieve alone, can be greatly facilitated by a common body. Cooperative structure can create the opportunity for business users to use not only their own data, but a much larger database, as well as to acquire and maintain expensive analytical capacities collectively. Such cooperation in data generation and analytics can be used to improve the competitiveness of SMEs vis-à-vis larger competitors with more market power, as well as data access and analysis capabilities.

Based on the results of the business user survey conducted for this study, four-fifths of respondents (81%) indicated that their business collects some data itself (i.e. collects at least one data type). The most prevalent data type collected by them is identification details of own customers (57% of enterprises collect this data), followed by business performance data (55%) and analysis of markets trends/developments (55%), see Figure 15.

FIGURE 15. DATA COLLECTION BY BUSINESSES THEMSELVES, BY DATA TYPE

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>Yes</th>
<th>No</th>
<th>Don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification details of your customers on the platform</td>
<td>57%</td>
<td>33%</td>
<td>9%</td>
</tr>
<tr>
<td>Identification details of other users on the platform</td>
<td>42%</td>
<td>47%</td>
<td>11%</td>
</tr>
<tr>
<td>Your customers’ behaviour on the platform</td>
<td>49%</td>
<td>40%</td>
<td>11%</td>
</tr>
<tr>
<td>Other customers’ behaviour on the platform</td>
<td>39%</td>
<td>49%</td>
<td>12%</td>
</tr>
<tr>
<td>Business identification of other business users on the platform</td>
<td>40%</td>
<td>48%</td>
<td>13%</td>
</tr>
<tr>
<td>Your business performance</td>
<td>55%</td>
<td>34%</td>
<td>11%</td>
</tr>
<tr>
<td>Data from individual transactions carried out</td>
<td>48%</td>
<td>40%</td>
<td>13%</td>
</tr>
<tr>
<td>Analysis of market trends/developments</td>
<td>51%</td>
<td>36%</td>
<td>12%</td>
</tr>
</tbody>
</table>

Source: PPMI; n=1667.

---


\(^\text{155}\) Fairmondo.
However, very few of the interviewed business users said they regularly collect data themselves to use it in their analyses and business decisions. The notable exception is app developers: as described in Section 3.2.2, they not only receive a significant amount of business-relevant data from the app stores, but also collect a lot of the needed information through their own applications. Meanwhile, in e-commerce, autonomous data collection by businesses mostly concerns business data which is not marketplace-specific, such as their business financials, advertising and product testing among target groups. In addition to this, some of the marketplaces allow tools or integrations allowing to automate this additional data collection for businesses. For instance, eBay and Etsy allow integrations with Google Analytics. In accommodation/hospitality segment, only large multi-city hotels seem to engage in active collection and analysis of data besides financials, such as customer details, customer feedback and trends in tourism. These large players tend to employ sophisticated automated data collection and analysis methodologies. Nonetheless, this mostly concerns the performance of their own company. Collection of competitor data is a challenge for individual business users, and they rely on specialised providers for this.

Meanwhile, most of the interviewees - especially in the e-commerce sector - reported that they use the services of third-party data and analytics providers (i.e. data brokers). For many of them these data were essential in developing new products, managing business and maintaining their business model more generally. Given that most competing business users are using the same data tools, others are also encouraged to use them too: data advantage means competitive advantage in the platform economy. The rather wide use of data companies is corroborated by the survey results. A third of surveyed companies (33%; 546) reported that they use third-party sources (data brokers) to access data (Figure 16).

![Figure 16. Use of Third-Party Sources to Access Any Types of Data](image)

Source: PPMI; n=1,666.

The following chapter, therefore, investigates these data companies, the services they provide and their value to platform business users in more detail.
6. Data companies

The importance of data companies is increasing. To illustrate, according to PwC estimations, data companies earned USD 21 billion in 2018\(^{156}\). Some authors even claim that the assumption that online platform giants Google or Facebook are the players that know every single detail on individuals is highly misleading\(^{157}\). Compared to the platforms, data brokers are not only equipped with larger datasets from online and offline sources, but also possess advanced and diversified techniques to refine and mine the data. They create a whole ecosystem of connected links on competitors, customers, market trends, products, services, and so on\(^{158}\).

These advantages attract not only smaller business players, but also big online platform economy players like Google, eBay, Amazon, Microsoft, AliExpress to actively purchase data brokers’ services\(^{159}\). For example, the study on partnerships between Facebook and data brokers has discovered that a large percentage of Facebook accounts is linked to the data broker information\(^{160}\). In 2017, it launched ‘Partner categories’ programme with the goal to establish a stronger relationship between Facebook and data brokers, such as Acxiom and Epsilon. Through the partnership, Facebook accessed the lists of people compiled by the data brokers, which enabled the campaign of targeted ads\(^{161}\). Despite terminating the programme in early 2018 as a part of a security reform, Facebook is one of the prominent examples in the data exchange between data brokers and platforms. Microsoft, eBay, Rakuten, Google are also known to use market intelligence tools such as SimilarWeb, Builtwith and SimilarTech to uncover the market dynamics and the competitive landscape. Generally, while platform data comes mostly from their own online activities, some data companies allow to put this into perspective of other online players and offline domain\(^{162}\).

As it was presented in the previous chapters, the value of data broker services can be especially significant for business users of online platforms. In fact, the evidence collected in the preparation of this paper shows that the data broker industry has become a key non-platform source for platform business users to access customised and relevant data\(^{163}\) that their platforms often do not provide (whether they have it or not).

Such data, analytics and insight providers are an important part of large ecosystems (or very broad networks\(^{164}\)) of companies evolving around the largest online platforms and providing various services


\(^{157}\) Mirani, L., & Nisen, M. (2014). The nine companies that know more about you than Google or Facebook. Quartz. Retrieved from [https://qz.com/213900/these-nine-companies-that-know-more-about-you-than-google-or-facebook/](https://qz.com/213900/these-nine-companies-that-know-more-about-you-than-google-or-facebook/)


for their business users - from consultancy, accounting and copyrighting services, to conferences and seminars, as well as specific data, analytics and insights aimed at helping business users to increase their market shares and profits. For example, some sources estimate hundreds of “trusted” companies offering solutions for selling on Amazon\textsuperscript{165}, with a notable category of them focused on data and analytics\textsuperscript{166}. The actual numbers, including smaller and more specialised providers, as well as companies not endorsed by the platforms, could be significantly higher. Similar clusters of service companies can be observed around all major multi-sided online platforms, including those examined in Chapter 4.

Given the large number of relevant providers in such ecosystems (like Unicorn Smasher, ZIK Analytics, Marmalead, Amalyze, Seller Board, Seller Labs, Helium10, AMZScout, etc., in the e-commerce and Appfollow, Appsflyer, Apptica, Flurry, App Census, etc., in the apps and software sector), we focussed our analysis to the key players serving the three types of business users analysed in the previous parts of this paper.

To identify and select data brokers for this analysis, we first searched for the most common and talked about data companies on business user online forums, such as Webretailer, Airhostforum, Reddit, Amazon Seller, Apple Developer Forum, etc. After this initial search, we identified the most relevant data brokers by focusing on their customers, usage, and outreach. Ultimately, we identified and analysed 15 data companies, serving the specific data needs of the three types of platform business users investigated this paper, and two data brokers covering multiple sectors of the platform economy and specific platforms analysed in this paper:

- App Annie, App figures, Apptopia, Mobile Action, Sensor Tower, serving the business users of application stores;
- Algopix, Jungle Scout, Sellics, Teikametrics, Terapeak for the ecommerce,
- AirDNA, Beyond Pricing, Uplisting, Wheelhouse, Skift for the accommodation/hospitality sector business users,
- Similar Web and Zirra for multiple sectors.

Although targeting businesses from different sectors and users of different platforms, all the business and monetisation models of selected companies depend on the access, quality and volume of data\textsuperscript{167}. They extract data from both businesses and consumers from various online (and sometimes offline) sources, and then aggregate for further analysis and development of insight products\textsuperscript{168}, tailored to platform business user needs. Their basic value proposition is comprehensive market intelligence based on the data in their possession and driven by advanced data analytics solutions\textsuperscript{169}. These activities, in turn, require an exceptional knowledge of data sources, collection methods and levels of innovation in analysis. The bigger, better and more diverse the accessible data is, in more services they can apply the data and increase their value to consumers. However, in all these processes they usually strongly depend on the very platforms whose business users they serve, and the data that these platforms make available in one way or another.

The table overleaf overviews the main characteristics, such as revenue, ownership status, headquarters (HQ), the presence of branches in Europe, the presence of branches in other regions, overall number of users (clients), the main names of business users and platforms that are clients, of data brokers analysed in this paper. Despite the fact that the majority of data companies (in our sample and overall) are located in the US, internet and branches in other regions enable a worldwide outreach, and they are widely used by European businesses. The report published by Transparency Market Research forecasts that North America will be the most attractive market for data brokers until 2026\textsuperscript{170}. Europe and Asia Pacific, in the meantime, will remain as other key markets for the data brokers\textsuperscript{171}. The main clients of investigated data brokers are business users of online platforms. However, during in-depth interviews some data brokers were not willing to share more specific information on their clients, especially names of platforms.


TABLE 9. DATA BROKERS’ OWNERSHIP STATUS AND LOCATIONS

<table>
<thead>
<tr>
<th>REVENUE</th>
<th>OWNERSHIP*</th>
<th>HQ</th>
<th>BRANCHES: EU</th>
<th>BRANCHES: OTHER REGIONS</th>
<th>OVERALL NR OF USERS</th>
<th>NOTABLE USERS</th>
<th>SPECIFIC PLATFORMS/ SECTORS COVERED</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>APP DEVELOPERS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>App Annie</td>
<td>$100M</td>
<td>App Annie</td>
<td>USA</td>
<td>Germany, United Kingdom, France, Netherlands</td>
<td>Canada, Japan, Singapore, China, Korea</td>
<td>Over one million worldwide</td>
<td>Visa, EPIC games, SoftBank, Tencent, Google, LinkedIn, Pinterest, Alibaba, Amazon App store</td>
</tr>
<tr>
<td>App figures</td>
<td>$3.8M</td>
<td>App figures</td>
<td>USA</td>
<td>No</td>
<td>No</td>
<td>More than 150,000 companies</td>
<td>Trello, Capcom, Snapchat, OneSignal, Microsoft, Airbnb</td>
</tr>
<tr>
<td>Apptopia</td>
<td>$4.2M</td>
<td>Apptopia</td>
<td>USA</td>
<td>No</td>
<td>No</td>
<td>NA</td>
<td>Spotify, Mattel, Animoca, Universal, RedBull, Uber, Google, Facebook, Google, Twitter, Uber, Pinterest</td>
</tr>
<tr>
<td>Mobile Action</td>
<td>$7.5M</td>
<td>Mobile Action</td>
<td>USA</td>
<td>No</td>
<td>No</td>
<td>Over 160,000 customers in over 65 countries</td>
<td>Adobe, Letgo, Playtika, Topps, eBay</td>
</tr>
<tr>
<td>Sensor Tower</td>
<td>$15M</td>
<td>Sensor Tower</td>
<td>USA</td>
<td>United Kingdom, China, Korea</td>
<td>NA</td>
<td>Zynga, NetEase Games, Funplus, Adobe, Reddit</td>
<td>Apple App Store, Google Play</td>
</tr>
<tr>
<td><strong>E-COMMERCE SELLERS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Algopix</td>
<td>$2.7M</td>
<td>Algopix</td>
<td>Israel/ USA</td>
<td>No</td>
<td>Israel</td>
<td>Over 65,000 worldwide</td>
<td>Spreetail, Sweetwater, Pulse commerce, Acenda, Blinq</td>
</tr>
<tr>
<td>Jungle Scout</td>
<td>$24.4M</td>
<td>Jungle Scout</td>
<td>USA</td>
<td>No</td>
<td>Canada, China</td>
<td>Over 200,000 worldwide</td>
<td>Entrepreneurs</td>
</tr>
<tr>
<td>Sellics</td>
<td>$9.6M</td>
<td>Sellics</td>
<td>German y</td>
<td>NA</td>
<td>USA</td>
<td>Thousands of brands and sellers</td>
<td>Spectrum, ERIMA, ABG, Roland, Bosch, Mammut, BRITA</td>
</tr>
<tr>
<td>Terapeak</td>
<td>$12M</td>
<td>eBay</td>
<td>USA</td>
<td>No</td>
<td>Canada</td>
<td>Millions of sellers</td>
<td>Hoopswagg, Bikewagon, WatchWarehouse</td>
</tr>
<tr>
<td>Teikametrics</td>
<td>$18.5M</td>
<td>Teikametrics</td>
<td>USA</td>
<td>No</td>
<td>No</td>
<td>NA</td>
<td>Razer, The Vitamin Shoppe, Swanson, Two Rivers Coffee</td>
</tr>
</tbody>
</table>

Information from.
<table>
<thead>
<tr>
<th>ACCOMMODATION SERVICE PROVIDERS</th>
<th>REVENUE</th>
<th>OWNERSHIP*</th>
<th>HQ</th>
<th>BRANCHES: EU</th>
<th>BRANCHES: OTHER REGIONS</th>
<th>OVERALL NR OF USERS</th>
<th>NOTABLE USERS</th>
<th>SPECIFIC PLATFORMS/ SECTORS COVERED</th>
</tr>
</thead>
<tbody>
<tr>
<td>AirDNA</td>
<td>$6M</td>
<td>AirDNA</td>
<td>USA</td>
<td>Spain</td>
<td>No</td>
<td>Over 10,000 active subscribers in over 80 countries</td>
<td>Hostmaker, Pass the Keys, Guestable, BookingPal</td>
<td>Airbnb, Vrbo, Booking.com, HomeAway</td>
</tr>
<tr>
<td>Beyond Pricing</td>
<td>$3.2M</td>
<td>Beyond Pricing</td>
<td>USA</td>
<td>No</td>
<td>No</td>
<td>Users in more than 7,000 cities</td>
<td>Casiola, Vacations Perfected, Sunset Properties</td>
<td>Airbnb, Vrbo</td>
</tr>
<tr>
<td>Uplisting</td>
<td>&lt;$1M</td>
<td>Uplisting</td>
<td>USA</td>
<td>United Kingdom</td>
<td>No</td>
<td>Over 14,000 successful properties</td>
<td>Airmanaged</td>
<td>Airbnb, Booking.com, Trivago, Vrbo, HomeAway, Expedia</td>
</tr>
<tr>
<td>Wheelhouse</td>
<td>&lt;$1M</td>
<td>Wheelhouse</td>
<td>USA</td>
<td>United Kingdom</td>
<td>No</td>
<td>Thousands of hosts and owners</td>
<td>AirbnbSecrets, Blanket Homes</td>
<td>Airbnb, TripAdvisor</td>
</tr>
<tr>
<td>Skift</td>
<td>$38.3M</td>
<td>Skift</td>
<td>USA</td>
<td>No</td>
<td>No</td>
<td>NA</td>
<td>Best Western, Hilton, Airbus, Hyatt, Air Canada, Google, Booking.com, Airbnb, TripAdvisor</td>
<td>Accommodation/hospitality/tourism</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OTHER</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SimilarWeb</td>
<td>$30M</td>
<td>SimilarWeb</td>
<td>Israel</td>
<td>United Kingdom</td>
<td>USA</td>
<td>From over 90 countries worldwide</td>
<td>Booking.com, Ryanair, Deloitte, P&amp;G, Adidas, Samsung, Pepsico, eBay, Google, Rakuten, Alibaba, Ryanair, Skyscanner, Booking.com</td>
<td>Retail, travel, consumer goods, publishing, ad tech, consumer finance</td>
</tr>
<tr>
<td>Zirra</td>
<td>$6.5M</td>
<td>Zirra</td>
<td>Israel</td>
<td>No</td>
<td>No</td>
<td>NA</td>
<td>Deutsche Telekom Capital Partners, Silverlake, Microsoft, AOL</td>
<td>Financial, investments, trading</td>
</tr>
</tbody>
</table>

Source: PitchBook, Crunchbase, Zoominfo, and data broker websites.
In the following sections, we use the selected sample of data brokers to explain the data processes within those companies, while also answering key research questions: first, what data do data brokers provide to business users on the three online platform categories and, second, to what extent data markets compensate for the data shortage, arising because of platform’s refusal to share this data. In the analysis, we also focus on the data brokers’ core features, such as the character of data they work with, sources and collection methods, management, and privacy and GDPR.

6.1. Types of data supplied by data brokers

Generally, as mentioned, some data brokers (including those in the analysed sample) prioritise particular industries and industry-specific data, while others grasp a wider field of information, based on which they develop services and products. On the one hand, the more varied and broader their data types are, the more opportunities to diversify their clients’ portfolio. On the other hand, collecting and building on industry-specific data types, allows to produce targeted in-depth insights, as well as builds credibility in a particular industry and value in the fast-growing and constantly changing data market.

Even though different industries require specific data collection and analysis, the data collected and provided by these players can be classified using the categories presented in Chapter 3:

- Customer demographics, e.g. age, gender, location, education, marital status, contact details. It basically builds profiles of business users’ customers.
- Business demographics, e.g. size, business model, products, location, contact details. Through this data type businesses can evaluate their own profile against competitors.
- Customer behaviour, e.g., clicks, browsing history, products and services purchased, allowing further insight into the current and future demand of specific products, trends, business niches.
- Business performance, e.g. ranks, daily visitors, search traffic, sales, revenue, web page activity, price changes, and similar aspects contributing to performance data. As mentioned earlier, data on competitors provides most value to individual businesses.
- Transaction information, e.g., transactions between businesses and consumers, payment details and methods; volume and value of the transactions.
- Market trends and analytics basically build on all the data types listed above to provide comprehensive overviews, leading to actionable insights and informed business decisions. Data brokers deliver detailed insights that may not accessible through online platforms or even business users’ own data analysis.

They combine anonymised and non-anonymised data that is bundled to deliver the product – market intelligence services. As illustrated in Table 10. Data types collected and supplied by data brokers, some differences in the data types provided to platform business users across the three sectors exist. E-commerce sellers access very similar data types across all companies, and very few data brokers provide information on customer demographics and behaviour, with the overall focus on own and competitor product metrics. App developers receive very similar information across all data companies, focusing on all the investigated data types. Accommodation service providers mainly access own performance metrics, where the main focus is pricing optimisation. Nonetheless, AirDNA also offers information on customer behaviour and demographics, as well as competitor performance data, while Skift offers research on the aggregate customer demographics, behaviour and market trends. Similar Web and Zirra, targeting broader user audiences, collect and provide information on performance metrics and market trends on a wide range of companies. Similar Web also offers information on customer demographics and behaviour, while Zirra offers information on business demographics and transaction.
All of the analysed data companies largely focus on three types of data - business demographics, business performance and market trends. This leads another important point about data collected by these companies - their data allows the businesses buying them to learn about their competition and get a **broader market overview**. As presented in Chapter 4, this is usually the kind of data, analytics and insights that many platform business users feel that they are missing.

Speaking about the level of data, as discussed in Chapter 3, the overviewed data brokers develop and supply **analytics and insights** products - that is, processed data to be used in business decisions, rather than raw data. Moreover, some of the data brokers offer advertising solutions by creating automations and promotions (e.g. Jungle Scout, App figures) - this would fall under the different category of our data classification, given its **actionable** character.

All the reviewed data companies offer this information in an interactive dashboard format, which includes graphs, pop-up boxes, tables, etc. A number of data brokers, such as AirDNA, Apptopia, Skift, Zirra, also produce reports filled with relevant information on competitor, performance, market, consumer trends. Some of them also offer browser integrations and API access.

Considering all the data, services and products that the analysed data companies provide, they can be located at the **left-hand end of the data value chain** (Figure 5). Whether they provide analytics and insights for individual sellers or broader overviews of market and competition, they mostly use the data that platforms share/allow to collect for anyone (but not everyone can use it for actionable insights). Their added value, in turn, is derived from their innovative tools and approaches to draw insights of this data. Sometimes these methodologies are so sophisticated that allow them to quite precisely estimate or extrapolate values of metrics that the platforms do not share.
### TABLE 10. DATA TYPES COLLECTED AND SUPPLIED BY DATA BROKERS

<table>
<thead>
<tr>
<th>E-COMMERCE SELLERS</th>
<th>APP DEVELOPERS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CUSTOMER</strong></td>
<td><strong>BUSINESS</strong></td>
</tr>
<tr>
<td>DEMOGRAPHICS</td>
<td>DEMOGRAPHICS</td>
</tr>
<tr>
<td>Algopix</td>
<td>NA</td>
</tr>
<tr>
<td>Jungle Scout</td>
<td>Consumer demographics</td>
</tr>
<tr>
<td>Sellics</td>
<td>NA</td>
</tr>
<tr>
<td>Teikametrics</td>
<td>Location</td>
</tr>
<tr>
<td>Terapeak</td>
<td>Location</td>
</tr>
<tr>
<td><strong>APP DEVELOPERS</strong></td>
<td></td>
</tr>
<tr>
<td>App Annie</td>
<td>Gender, age (aggregate form)</td>
</tr>
<tr>
<td>App figures</td>
<td>Consumer demographics, contacts</td>
</tr>
<tr>
<td>Apptopia</td>
<td>Age, gender (aggregate form)</td>
</tr>
<tr>
<td>Mobile Action</td>
<td>Consumer demographics</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Sensor Tower</td>
<td>Age, gender (aggregate form)</td>
</tr>
</tbody>
</table>

**ACCOMMODATION SERVICE PROVIDERS**

<table>
<thead>
<tr>
<th>AirDNA</th>
<th>Location, gender, age (aggregate form)</th>
<th>Company address, VAT number, country of operation</th>
<th>NA</th>
<th>Pricing, average daily rate, occupancy rate, revenue, ratings, demand/growth</th>
<th>Top properties, daily rate, occupancy, pricing, revenue, amenities, ratings</th>
<th>Currencies, reviews, booking time, amount paid, ratings</th>
<th>Market, sub-market, competitor, performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beyond Pricing</td>
<td>NA</td>
<td>Company address, VAT number, country of operation</td>
<td>NA</td>
<td>Pricing, bookings, demand, occupancy</td>
<td>NA</td>
<td>Reviews, ratings</td>
<td>Localised market, neighbourhood, seasonality</td>
</tr>
<tr>
<td>Uplisting</td>
<td>NA</td>
<td>Company address, VAT number, country of operation</td>
<td>NA</td>
<td>Pricing, average daily rate, nights sold, average length of stay</td>
<td>NA</td>
<td>Reviews, ratings</td>
<td>Performance</td>
</tr>
<tr>
<td>Wheelhouse</td>
<td>NA</td>
<td>Company address, VAT number, country of operation</td>
<td>NA</td>
<td>Pricing, average occupancy, average night rate</td>
<td>NA</td>
<td>Reviews, ratings</td>
<td>Performance</td>
</tr>
<tr>
<td>Skift</td>
<td>Customer demographics</td>
<td>NA</td>
<td>Customer habits and behaviour</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Market</td>
</tr>
</tbody>
</table>

**OTHER**

<table>
<thead>
<tr>
<th>Similar Web</th>
<th>Gender (aggregate form)</th>
<th>NA</th>
<th>Audience interests, traffic sources, clicks, browsing history, other products or services purchased on the platform</th>
<th>Website, industry and app ranks, total visits, traffic per country</th>
<th>Popular pages, ad campaigns</th>
<th>NA</th>
<th>Consumer, business, performance, keyword</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zirra</td>
<td>NA</td>
<td>Company address, VAT number, country of operation</td>
<td>NA</td>
<td>Web traffic, employee growth and venture, business models, key risks, achievements, employee count</td>
<td>Overall financial information, reviews</td>
<td>Company, stock</td>
<td>NA</td>
</tr>
</tbody>
</table>

Source: Data broker websites.
Data brokers serving app developers, e-commerce sellers and accommodation service providers focus on their markets by providing services that reflect their specific customers’ needs - depending on a business segment or a specific platform that they operate in. We further discuss these in more detail by business segment.

6.1.1. E-commerce

The desk research and interview data indicate that platforms in the e-commerce segment are the focus of large networks of ancillary service and Software-as-a-Service (SaaS) companies, engaging in everything from accounting, copywriting and advertising, to development of sophisticated data-powered tools for sellers. The platform’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.

Data brokers specialising in e-commerce, mostly collect and analyse data on business demographics, market, consumer, product trends, and similar data, on third-party sellers (i.e. business users) on specific platforms. Fewer data brokers provide transaction information and customer demographics, and basically none offer detailed customer data.

Some SaaS companies, such as Teikametrics for Amazon business users, provide sellers with data-powered tools to build successful brands by discovering new keyword opportunities, automating keyword bids or eliminating advertising which does not generate the satisfactory levels of conversion[173]. 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).

Nonetheless, as outlined in Chapter 3, e-commerce sellers are especially interested to know about the business performance of their competitors, as well as overall market trends. 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 data brokers 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 - even though the platform does not share this data itself. The mentioned providers allow to acquire substantial data on the competitors, and the estimations are said to be “spookily accurate”[174]. We provide more details on how data brokers collect such data in the following section.

Also, Amazon, eBay and Etsy sellers note that compared to the analytics and insights provided by platforms, third-party providers are faster and more comprehensive, especially useful if the sellers need to quickly understand how their business are doing. On the other hand, 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.

The use of such solutions is very widespread in the e-commerce segment. As several interviewees reported, as far as they knew, most sellers (especially the larger ones) use such services. As an interviewee explained, if your competitors use it, you must use it too to be successful; data disadvantage can be very costly in this business.

6.1.2. Software development

The analysed data companies specialising in serving software developers data needs, mostly focus on mobile economy metrics. 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 historical market, competitor or performance trends. Some of the reviewed companies also collect consumer demographics and behaviour. In an aggregated form, these result into products such as app analytics, app optimisation, market research, app information, providing of a good overview of the specific app markets and competition.

Such 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 and SensorTower, fill this gap by presenting comprehensive app market data, including performance of specific apps and markets, as their main selling point. Interviewed app developers mentioned that they use the sources together with the app store data extensively.

In addition to this, just as in case of other segments of business users, 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\(^{175}\) provide solutions based on the developers data, aimed at profiling users to increase their engagement and optimise conversion. App Figures, which also provides insights into market and competitors, offers data products to track the performance of individual apps of the app developer in one place.

Generally, app developers tend to use a number of data and analytics services for developing their products and running the business. The interviewees mentioned around two to four different third-party data/analytics/insights providers whose solutions they use.

Another segment of data providers, focused on app developer needs, specialises in data feeding into the apps themselves. 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 built sophisticated tools that allow their customers to access the data in a controlled way. 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. Players like Amazon and Google also act as data providers and are launching a 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 to avoid digital hoarding.

6.1.3. Accommodation and hospitality

Data providers specialising in tourism, accommodation and hospitality mostly focus on rental listings’ data on various OTAs. They provide information 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. For example, AirDNA 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. Its wide coverage helps understand own market and other market trends from different angles.

Importantly, companies such as AirDNA provide estimated and extrapolated insights into the data that the OTAs do not share. To illustrate, in late 2015 Airbnb stopped providing the overall real-time reservation data. AirDNA, in turn, uses an algorithm based on 16 indicators picked up in historical data to determine the reservation status for each listing. They argue that their algorithm has an error margin of only 5% and therefore provides very precise information, which can be further used by accommodation providers to estimate the future demand and set appropriate prices.

Other providers, such as Phocuswright and Skift, gather 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 specific information on business on various online platforms.

Nonetheless, according to the interviews and desk research into the specificities of hotel business, the use of data tools in the accommodation sector is much less widespread, as compared to other platform business segments. This is related, at least in part, to the fact that vast majority of European properties listed on OTAs are small businesses, whose market and competition is very strictly geographically limited.

6.1.4. Non sector-specific data providers

Some platform business users across all platforms and sectors also use third-party data providers to track the more general and non-platform-specific developments in online markets. Some of them allow to better understand the individual business user’s performance. For instance, Google Analytics is used widely, especially by the business users of platforms that allow Google Analytics integrations (e.g., Etsy and others). It provides various metrics on the business user listing pages, as well as allows to check the performance of their marketing and content. What is more, according to an accommodation sector interviewee, the added value of Google Analytics is information on customer age, clicks on a website, how long customers spend on a website, why and when they leave, and so on - something that most platforms do not provide.

Other providers that do not focus on specific platforms also allow to take a look into the performance of other businesses. For example, ComScore was mentioned by an interviewed business user as an important source of marketing data. Among other insights, it provides metrics on traffic of specific websites, allowing to understand specific markets (e.g., travel) and competitors better. SimilarWeb offers its customers

---

177 https://www.phocuswright.com/
178 https://skift.com/
information on their clients’ and competitors’ website traffic volumes, referral sources, and website "stickiness" (e.g., time on site, page views, bounce rate). Zirra specialises in ratings and analytics of private companies. These providers, however, are more interesting for large businesses in competitive markets.

### 6.2. Data sources and collection methods

Given that the access to data points unavailable elsewhere and generation of unique insights is the basis of data brokers’ value proposition for their customers, they have become very sophisticated in data acquisition and analysis methods. Data brokers use a wide array of similar online and offline sources (including commercial and independent entities) to acquire the data\(^ {179}\) as well as a number of highly advanced technical methods. While some of this information is easily accessible with the right tools, some requires significant financial resources. We further describe the main sources as well as data collection methods that the data companies exploit in more detail. For clarity, the box below also describes key definitions of technical data collection method terms used further in this paper.

**BOX 1. KEY DEFINITIONS OF TECHNICAL DATA COLLECTION METHODS**

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artificial Intelligence</td>
<td>The ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings.</td>
</tr>
<tr>
<td>Cookies</td>
<td>An electronic monitoring or a file or part of a file saved to a web user's hard disk by a website.</td>
</tr>
<tr>
<td>Digital fingerprinting</td>
<td>A fingerprinting algorithm is one that reduces a larger data set to a very small data set, sometimes called a bit string, to promote efficient identification and search protocols.</td>
</tr>
<tr>
<td>Proprietary importing</td>
<td>These are algorithms that allow to remove and consolidate duplicates.</td>
</tr>
<tr>
<td>Web beacons</td>
<td>A tiny graphic image placed on a website or in an email that is used to monitor the behaviour of the user visiting the Web site or sending the email. It is often used in combination with cookies.</td>
</tr>
<tr>
<td>Web crawlers</td>
<td>An Internet bot which helps in Web indexing. They crawl one page at a time through a website until all pages have been indexed. Web crawlers help in collecting information about a website and the links related to them, and also help in validating the HTML code and hyperlinks.</td>
</tr>
<tr>
<td>Web indexing</td>
<td>The processing of the information gathered by the bots from their crawling activities.</td>
</tr>
<tr>
<td>Web scraping</td>
<td>A form of data mining, which is done through a software that simulates human web surfing to collect specific bits of information from different websites.</td>
</tr>
<tr>
<td>Web tracking</td>
<td>The act of archiving existing websites and tracking changes to the website over time.</td>
</tr>
</tbody>
</table>

Source: Encyclopædia Britannica & Techopedia.

**Online sources**, unsurprisingly, are key to the data companies serving platforms business users. These cover automated collection of all the publicly available data on the internet, including social media, web browsing, blogs, media reports, websites, application stores, and so on. This type of information is mostly gathered from websites on submitted job applications, cookies, ‘contact us’ forms, and is a very popular data source\(^ {180}\). Every data broker in our sample uses internet data as the main source of information.

---


Generally, the online data collection methods that data brokers employ depend on the specific information source. Many data brokers use data science- and technology-powered tools such as AI, machine learning, proprietary importing infrastructure, web crawlers, website tracking, internet indexing. For example, data companies track cookies, which in addition to easing the user experience online, such as saved login details or search history, is a great tool to collect information on individuals’ online behaviour\textsuperscript{181}. Other means may involve digital fingerprints that trace information from web browsers, computers, or smartphones.

Furthermore, for data providers focusing on specific platforms, these \textbf{platforms are the main source of data} used in further analytics. Below we provide several examples of the most widely used automated online data collection methods to collect platform-specific information.

\begin{itemize}
  \item \textbf{Web scraping}. A large number of third-party data companies engage in scraping of platform websites. Amazon representatives revealed in an interview that around 35\% of the traffic to Amazon’s stores comes from web scraping tools. The online retailers and data 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 is not completely compliant with the platform’s terms and conditions. Given the precision of insight they have, it is likely that the top data providers to Amazon sellers (including Jungle Scout) should have a more direct contact with the platform itself. Many other platforms do not permit scraping for commercial purposes, but that does not prevent data companies from collecting data this way.

  \item \textbf{Crowdsourcing business user account data}. Some third-party data providers use the platform-generated data of their customers (i.e. platform business users). More specifically, as described in Chapter 4, each seller on e-commerce marketplaces gets access to their own business insights, as well as some insights into their competitors and the broader market overviews. Some third-party data providers ask online sellers to share their marketplace information, and then link this data of thousands of users to draw market insights. For example, Jungle Scout collects data from a large number of actual sellers (over 225,000) who have opted in to share their sales information with them\textsuperscript{182}. When merged with the data gathered by scraping the platform’s front-end (e.g., Best Seller rank on Amazon), this can become basis of quite precise estimations and extrapolations. Similarly, if AirDNA users wish to receive performance analytics, they will be asked to upload their Airbnb host IDs. After doing this, they can see their performance trends, comparative and financial analysis on all vacation rental listings. Other data brokers, such as Beyond Pricing, Apptopia (with 300,000 developers opted-in), Wheelhouse, AppAnnie (with a million developers opted-in) and others use such technique as an important data source.

  \item \textbf{Processing and analysing data provided by platforms}. Some platforms tend to provide a lot of raw data. As many business users simply do not have these large data science and IT departments that could work with it, third-party data providers step in to fill this gap. For example, Terapeak has had this role for eBay sellers (both before and after its acquisition by eBay). However, relying on these sources can be risky for the data companies. First, their business model fails if the platform starts providing the same data or analytics itself. For instance, Allegro, a Polish e-commerce platform freely shares most of the data on the activities on their platform via an API. It has been long used by

\end{itemize}


#:~:target=Text=We%20gather%20it%20from%20a,their%20sales%20info%20with%20us.&targetText=This%20means%20that%20we%20analyze%20and
%20test%20it%20daily.
an insight provider focusing on this data. However, recently the platform itself made significant investments into analytics, to develop new data products (Allegro Statistics) that are now provided to its sellers, endangering the business of third-party analytics providers. 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. Similarly to the AirDNA approach presented above, these third-party providers ultimately adapted to these changes by leveraging precise historical data from the API together with sophisticated means of forecasting and estimating\textsuperscript{183}. However, their product is not of the same 100% level of accuracy anymore.

Overall, however, the relationships between specific platforms and data brokers that collect data on them vary from case to case. Some of the interviewed platforms did not comment on their relationships or the validity of data that the third-party data companies provide. Several interviewed data companies, in turn, claimed that they have good relationships with the platforms they focus on, while others did not comment on this. Reviewed online sources also show that platforms invest in fighting illegal data collection. For example, Amazon banned several entities that were acquiring the user data from the platform in violation of its policies\textsuperscript{184}. Nonetheless, the interviewed representative of Amazon presented the current landscape of third-party providers, as helping to meet the sellers’ data needs.

Further, non-automated (or semi-automated) data collection methods are also applied by data brokers, especially those that do not focus on specific platforms but rather develop broader market and consumer insights. Some of the main ones are the following:

— Data acquisition from commercial and marketing entities/ third parties. Private contractual agreements between data brokers and private entities are quite common when acquiring commercial data, which includes but is not limited to consumer demographics, consumer behaviour, transaction information, etc\textsuperscript{185}. The contractual agreements also include a range of provisions and warranties that secure a legal collection and distribution of their data\textsuperscript{186}. Some specific examples of such entities selling data to data brokers include:

- Platform operators, including those in active in advertising (e.g. Facebook, Google, Twitter, etc.) and the platforms of interest themselves.
- Retailers and other commercial companies that sell information about their customers, e.g. purchasing habits, registration information, transactions, etc. Specific names of companies are not disclosed since this information is private and not shared by data companies because

of claimed competition issues\textsuperscript{187}. However, we discovered that several data brokers analysed in this paper do use third parties to receive information about their business users\textsuperscript{188}.

- Financial institutions, such as banks, credit unions, brokerage services, insurers, employers.
- Other data brokers specialising on specific sectors/platforms.

More “traditional” data collection methods, such as \textit{manual sourcing, desk research, direct phone interviewing, customer panels, and use of official statistics}. For instance, AppAnnie, among other sources, use data of consumer panels. ComScore maintains a group of users who have monitoring software installed on their computers. In exchange for joining the ComScore research panels, users are presented with various incentives, such as free computer security software, Internet data storage, virus scanning and chances to win cash or prizes. Meanwhile, Skift, in addition to using automated data collection, also employs the old-fashioned methods of manual research, expert interviews and publicly available information.

The overviewed sample of data brokers use a combination of data collection methods (see Table 11). This way they can expand their outreach and competitiveness. Generally, the main source is business user data, where business users integrate their accounts with the data brokers’ platform. Data brokers mostly employ data science and technology-driven tools to gather data from online sources.

\begin{center}
\textbf{TABLE 11. DATA COLLECTION SOURCES AND METHODS OF ANALYSED DATA BROKERS}
\end{center}

<table>
<thead>
<tr>
<th>E-COMMERCE SELLERS</th>
<th>ARTIFICIAL INTELLIGENCE</th>
<th>MACHINE LEARNING</th>
<th>WEB SCRAPING</th>
<th>BUSINESS USER DATA</th>
<th>PLATFORM DATA</th>
<th>DATA FROM THIRD PARTIES</th>
<th>MANUAL RESEARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algopix</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jungle Scout</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sellics</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teikametrics</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terapeak</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APP DEVELOPERS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>App Annie</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>App figures</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apptopia</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobile Action</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensor Tower</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACCOMMODATION SERVICE PROVIDERS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AirDNA</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beyond Pricing</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uplisting</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wheelhouse</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skift</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OTHER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Similar Web</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zirra</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Data broker websites


6.3. Data management and protection

Data brokers’ objective is to add value to customers, by providing needed information at high levels of quality, validity, volume and granularity. We further describe how data brokers manage and structure their data on internet users.

A general practice in data management is to create a unique ID for every entity (individual, company, product etc.), which serves as the main identifier in the data brokers’ system. Some of the main data elements that create a unique ID are personal data, such as name, address, phone number, email, device ID. Those data elements can be of finite or free form, which influences data structuring. The free-form data elements (e.g. name and address) are harder to assign as an ID and can lead to matching errors (because of more symbols used and hence more possible variation in how they are entered in different datasets), whereas finite elements (e.g. email and phone number) create straightforward matching.

Once the unique IDs are created, they are assigned values and characteristics. One of the most common ways to build datasets is ID synchronisation: matching IDs assigned to internet users by one entity (data broker, platform, website, etc.), to IDs assigned to other entity. Website visits, cookies, media, apps are a part of the ID syncing. For example, data broker has assigned a visitor ID 1234. Another platform knows this visitor by ID 4321. The data broker maps these IDs together during the synchronization process. The results add new data points to what they know about the internet users (be it app end-user or e-commerce buyers, etc.).

Nonetheless, the ID syncing is only one side of the coin. The accumulation of information creates the need for big data analytics with correct linkages of multiple data points. Data mining helps to classify data into different categories, and therefore create a structured dataset suitable to data buyers’ needs. It also lets data brokers profile internet users. To ensure correct linking between the data and unique IDs, data brokers may also perform data enrichment, appending and enhancement. Data brokers enrich datasets by appending information to the IDs from the data buyer’s list. Data enhancement is a tool that ensures every data point is relevant with a maximised value. This structures the data and allows for detailed insights on companies, individuals and households, such as demographic, household, financial, lifestyle, behaviour and any other information that might be of interest to the data buyer. This ultimately results in comprehensive datasets, enabling the data scientists to make seamless and traceable inquiries, and quick extraction of specific data points.

During our interviews, none of the interviewed data brokers provided a thorough answer on data management techniques. Nonetheless, the interviewees revealed that data brokers regularly update their models to ensure accuracy and appropriateness of their data management. For instance, Apptopia shared

---

that they update their models at least once a quarter. Also, all data brokers claim to anonymise and aggregate the final data product that they offer to clients.

At a first glance, therefore, data brokers seem compliant with data protection and privacy laws, especially GDPR in the EU. However, the huge amounts of data that data brokers collect, store, possibly re-personalise and disseminate do pose privacy concerns. With today’s technological capabilities and vast data sharing networks, data brokers have the tools to continuously enrich, trade and share personal data. Most of the time individuals or companies are unaware of what information data brokers collect on them or even that they collect information at all. Due to this information asymmetry, the data broker industry is quite often characterised as opaque, non-transparent, arbitrary, biased, unfair and unaccountable. Their goal is to accumulate as much data as possible, especially personal data of individuals and businesses. This raises many concerns in media, policy and research communities.

Since 2018, the GDPR provides the general data protection framework. It requires companies that collect data to present instructions on cookies, opt-in/opt-outs and ask for explicit consent from users. Privacy policies, terms and conditions must provide information on data gathering and processing. Under this new regulation, data companies had to undergo massive changes to ensure the compliance of their data gathering and processing practices. Some of them had to restructure and drop some of their more controversial services. For example, Axiom sold off a part of its business that offered identity resolution services (LiveRamp), while some other companies withdrew their data products from the market.

The compliance of data companies has not always been ensured successfully. For example, a few months after the GDPR came into force, Privacy International filed a complaint against seven data brokers: Axiom, Oracle, Criteo, Quantcast, Tapad, Equifax, and Experian. The main argument was their incompatibility with the GDPR and failure to comply with data protection principles (such as acquiring consent, providing detailed and transparent information for the data subject access requests), exploiting data in unknown ways.

Nonetheless, personal data allowing to match and link the data from online and offline realms, remains paramount for data brokers. All data brokers analysed in this paper collect personally identifiable information (PII) and non-personally identifiable information. PII refers to any data that can be used to identify a specific individual or business. The data companies claim that they anonymise and aggregate information to such a level that does not allow personal identification. For this, some data brokers use advanced encryption tools to encrypt the sensitive and private information. However, according to some

---

202 https://liveramp.com/
accounts, it is questionable whether the security measures that data broker have in place can guarantee a secure data storage and transmission\textsuperscript{205}.

Currently, many data brokers collecting data on EU users apply complex analysis of trends rather than focus on individual level data\textsuperscript{205}. Nonetheless, in some cases the most basic metrics can be a credible resource of building individual profile or re-personalisation. In fact, the information that data brokers collect in internet users may still be sufficient to prevent proper protection of their data (see the box below). Moreover, a number of news articles exist on new leaks about data broker’s (illegally) stored data\textsuperscript{207}.

**BOX 2. ISSUES WITH DATA PROTECTION AND PRIVACY**

| The protection and privacy of personally identifiable information of individuals and businesses has become a contentious topic. Nonetheless, data brokers continue to thrive on data by collecting and processing a lot of it, because internet users continue to agree to terms and conditions and grant access to data collection and use for sometimes unknown purposes. Data brokers’ privacy statements are often long and confusing, and have complicated wording. Policies also tend to change quite often, leaving no time to keep up with updates and detect every new detail added to the privacy statement. One can also claim that most of the alleged openness and transparency is rather complicated. To be sufficiently aware of what is done with their data, individuals have to invest significant time, energy, and sometimes money to find how to file the data subject access request (DASR) or understand how exactly the data is being collected. When asking for the latter information, a number of companies request to register in order to apply for DASRs. We also tried to file DASR on MasterCard, Equifax, Epsilon, Quantcast websites. Proceeding with the request meant giving full name, email, contact number, birth date, address details; some company websites required to enable cookies. The increasing demand for personal data has created the ‘privacy paradox’. The term has become increasingly relevant in the 21\textsuperscript{st} century and refers to internet users that are concerned about their data security and eroding privacy, and yet, they are willing to give up personal data in exchange of accessing services (sometimes those can be “free” in exchange for data). Data brokers use this to get data from internet users, who give up their personal data without realising it. Long privacy statements, terms and conditions with ‘take it or leave it’ position, result in consequences on privacy and control over personal data processing. |

Analysis of the selected data brokers shows that in some cases the level of their compliance with GDPR is not clear. Some of them do not provide information on GDPR compliance. We gathered information from privacy policy statements of selected data brokers. The table below presents the findings, including privacy policy updates and information about the GDPR compliance measures.

\textsuperscript{205} For instance, Ruhaak, A. (2019). „Data Brokers Are Cruising for a Bruising.” Wired. Available at: https://www.wired.com/story/opinion-data-brokers-are-cruising-for-a-bruising/


### TABLE 12. ANALYSED DATA COMPANIES’ COMPLIANCE WITH GDPR

<table>
<thead>
<tr>
<th>Company</th>
<th>Privacy Policy Last Modified</th>
<th>Stated GDPR Compliance</th>
<th>Information on GDPR in the Privacy Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E-COMMERCE SELLERS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Algopix</td>
<td>6 June 2019</td>
<td>Yes</td>
<td>&quot;...Data subjects with respect to whose data GDPR applies, have rights under GDPR and local laws, including, in different circumstances, rights to data portability, rights to access data, rectify data, object to processing, and erase data...&quot;</td>
</tr>
<tr>
<td>Jungle Scout</td>
<td>3 February 2016</td>
<td>Unknown</td>
<td>NA</td>
</tr>
<tr>
<td>Selics</td>
<td>Unknown</td>
<td>Unknown</td>
<td>NA</td>
</tr>
<tr>
<td>Teikametrics</td>
<td>6 February 2014</td>
<td>Unknown</td>
<td>NA</td>
</tr>
<tr>
<td>Terapeak</td>
<td>27 November 2017</td>
<td>Unknown</td>
<td>NA</td>
</tr>
<tr>
<td><strong>APP DEVELOPERS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>App Annie</td>
<td>23 May 2018</td>
<td>Unknown</td>
<td>NA</td>
</tr>
<tr>
<td>App figures</td>
<td>24 May 2018</td>
<td>Unknown</td>
<td>NA</td>
</tr>
<tr>
<td>Apptopia</td>
<td>15 November 2018</td>
<td>Unknown</td>
<td>NA</td>
</tr>
<tr>
<td>Mobile Action</td>
<td>24 May 2018</td>
<td>Unknown</td>
<td>NA</td>
</tr>
<tr>
<td>Sensor Tower</td>
<td>15 June 2018</td>
<td>Unknown</td>
<td>NA</td>
</tr>
<tr>
<td><strong>ACCOMMODATION SERVICE PROVIDERS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| AirDNA                 | Unknown                      | Yes                    | "AirDNA has undergone a full audit under the new GDPR regulations, which came into force on May 25, 2018."
|                        |                              |                        | "...If you wish to be deleted entirely from the AirDNA customer database under the GDPR Article 17 “Right to Erasure” (May 2018), email us..." |
| Beyond Pricing         | 25 May 2018                  | Unknown                | NA                                       |
| Uplisting              | 2 March 2018                 | Unknown                | NA                                       |
| Wheelhouse             | 7 June 2018                  | Unknown                | NA                                       |
| Skift                  | 26 September 2018            | Yes                    | "...On this page, we’ll explain our approach to achieving GDPR-compliance, both for ourselves and for you..."
|                        |                              |                        | "...We’re committed to our users enjoying Skift to help them understand and prepare for the General Data Protection Regulation (”GDPR”). The GDPR is the most comprehensive EU data privacy law in decades..." |
| **OTHER**              |                              |                        |                                          |
| Similar Web            | 25 May 2018                  | Unknown                | NA                                       |
| Zirra                  | 2019                         | Unknown                | NA                                       |

Source: Data broker websites.

*For some data analytics companies that have updated their privacy policy around May 2018, but mention no specifics on the GDPR, one may assume that they took into account the GDPR. Alternatively, if data brokers last modified their privacy policy a few months prior to or after the GDPR enforcement, but mention no specifics in their privacy policy, it is either that those companies are compliant or they have other ways to deal with the law. Even though these are just assumptions, they are a good indicator of the lack of transparency and dubious approaches to data privacy within the EU, and invite to question further.*
This paper presents a detailed analysis of the platform business user data needs, their access to data and analytics generated by the platforms that they use, as well as the alternative means of gathering the needed information. Building on the diverse evidence from research literature, interviews, online sources and an online survey, the analysis provided insights into several topics related to data in the P2B relationships.

To begin with, so far, the research and policy discussion of business access to platform data have been rather unspecific. Several earlier studies have indicated the dissatisfaction of business users with the level of platform data access. However, there was little information on what data exactly is needed, for what purposes, and what are its potential costs. Chapter 3 of this paper took a deeper look into this topic by classifying the types and levels of data demanded by the platform business users from three segments of platform economy: e-commerce, hospitality and app distribution. Three general dimensions of data relevant to platform business users were distinguished:

- the type of data by object (customers, businesses, user behaviour, markets, transactions, etc.);
- whether the data is about an individual business which receives it, or other businesses on the platform (competition)/ whole marketplace.
- by the level of data processing and its value, from raw datasets to insights guiding business decision-making.

When asked about their data needs, businesses using online platforms do not significantly differentiate between the usefulness and relevance of data on customers, other businesses, user behaviour, markets and transactions, and report their interest in all these types of data. Similarly, most of them use (or would like to use) both, data on their individual performance, as well as data on their competitors and overall market trends. Meanwhile, what concerns the level of data processing, most firms do not have the capacity to deal with raw datasets but need data which is sufficiently processed and can provide actionable insights to keep pace with rapid market developments and improve their business performance.

Second, in Chapter 4, the paper investigated the data access provided by the four major platforms - Amazon, eBay, Google Play and Booking.com to their business users. The analyses showed that the kinds of data provided and not shared by these players are rather similar. To start with, all the analysed platforms have made notable investments in analytics capacities to draw insights from this data and provide them to the business users. Sharing of such data is beneficial for the platforms. Data access for business users, as well as expensive advanced analytics, are granted to the extent that it could generate more income for the platform as the business users make more transactions. Nonetheless, the platforms do neither share the raw big data on day-to-day activities, nor detailed data on customers and competitors. Generally, although they collect and analyse loads of data, only a fraction of this is provided to other players.

A notable share of business users report that they do not get sufficient access to data from platforms. Platforms, in turn, provide numerous justifications for not granting the access to data, including the protection of their business model, protection of their trade secrets, issues of interoperability, need for significant investments to make the data useful, lack of necessity to provide certain data, legal barriers and customer interests.
The refusal of platforms to share data can be especially problematic when a platform is vertically integrated and has significant market power. Unique and comprehensive datasets on all the firms and their customers operating in the marketplace can give a huge business advantage to the platform operators. If the platform is also a significant player in the market, its business users do not have much choice of switching marketplaces without serious losses.

Third, the analysis found that platform business users can get the data and insights they need from alternative sources, by either collecting and analysing it themselves or using services of third-party providers and data brokers.

In fact, analysis of the latter showed that all the major platforms have become the centre of large ecosystems of companies, providing various services to the business users operating on these platforms. Companies specialising in data and insights are significant players in these ecosystems, providing services to large numbers of business users. These data brokers fill some of the important data gaps, especially as concerns data on competition and actionable business insights.

However, the data needs of platform business users are often very specific, and differ depending on their analytical capacities, and especially on the platforms or marketplaces they use. Most of the information they need concerns specifically the platforms that they use. Data brokers do collect a lot of data. However, whether this data presents a full-fledged substitute for the data collected by specific marketplaces, app stores or OTAs and not shared with the business users, cannot be answered straightforwardly.

Importantly, most of the data these companies work with comes from the platforms themselves - either by scraping platform front-end or acquiring access to individual analytics from registered platform business users (who provide this access to the third parties to be able to use their data services). It seems that platforms virtually close the door (by not providing some data and insights) but leave a window open for data collection, and third-party providers exploit this. The data brokers step in with superior technical and analytical capacities, innovative tools and approaches to collect and develop the insights that businesses need. Whether the data companies provide insights for individual traders or broader overviews of market and competition, they mostly use the data that platforms do share or allow to collect for everyone. The added value of these companies is that business users themselves do not need to invest into any analytics or IT, but rather buy products tailored to their needs. Sometimes the methodologies applied are so sophisticated that allow them to quite precisely estimate or extrapolate the metrics that the platforms do not share. However, at the same time, the data companies remain highly dependent on the platforms that they analyse and on any changes in their data sharing policies.

The question whether the platforms should be obliged to share more data is complicated. While improving the access to data to many economic players, this may have negative effects on all sides in platform economy. To begin with, providing everyone with the same access to data may be detrimental to the business users themselves, whose financial, client and performance information would be disclosed to competitors. Secondly, the impact of data sharing should also be assessed from the consumer perspective. Platforms invest a lot in customer satisfaction, and data protection is part of the consumer experience. Consumers prefer to deal with centralised services rather than with multiple actors asking for consent, and they are not willing to be flooded with marketing messages. Further, mandatory data sharing may have a negative impact on platforms’ business and profits209. In such a case, platforms may even 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.

While this analytical paper presents an overview on the state of the art of the data access for platform business users, it also allows to raise new questions and hypotheses for further research. To begin with, the analysis revealed concrete issues in specific segments of the platform economy. Firstly, the most problematic cases of not sharing data concern vertically integrated platforms with significant market power, who can use data to their advantage. Secondly, this paper looked into three key sectors of the platform economy: e-commerce, app development and accommodation/hospitality. In order to get further insights on data access, further investigation into other sectors not covered in this paper is needed. Thirdly, further research would be useful into the legal and technical aspects of the platform data and sharing, including interoperability.
## List of interviewees

<table>
<thead>
<tr>
<th>#</th>
<th>ORGANISATION/ COMPANY</th>
<th>BUSINESS TYPE</th>
<th>INTERVIEW DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ONLINE PLATFORMS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>Amazon marketplace</td>
<td>E-commerce</td>
<td>03/12/2019</td>
</tr>
<tr>
<td>2.</td>
<td>Allegro</td>
<td>E-commerce</td>
<td>06/12/2019</td>
</tr>
<tr>
<td>3.</td>
<td>eBay</td>
<td>E-commerce</td>
<td>17/10/2019</td>
</tr>
<tr>
<td>4.</td>
<td>eMAG</td>
<td>E-commerce</td>
<td>31/10/2019</td>
</tr>
<tr>
<td>5.</td>
<td>Etsy</td>
<td>E-commerce</td>
<td>22/11/2019</td>
</tr>
<tr>
<td>6.</td>
<td>Real.de</td>
<td>E-commerce</td>
<td>16/11/2019</td>
</tr>
<tr>
<td>7.</td>
<td>Zalando marketplace</td>
<td>E-commerce</td>
<td>20/11/2019</td>
</tr>
<tr>
<td>8.</td>
<td>Booking.com</td>
<td>Accommodation/ hospitality</td>
<td>08/11/2019</td>
</tr>
<tr>
<td>9.</td>
<td>Expedia Group</td>
<td>Accommodation/ hospitality</td>
<td>19/11/2019</td>
</tr>
<tr>
<td>10.</td>
<td>TripAdvisor</td>
<td>Accommodation/ hospitality</td>
<td>13/11/2019</td>
</tr>
<tr>
<td>11.</td>
<td>Apple (App store, iOS, sdk)</td>
<td>Apps and software</td>
<td>30/10/2019</td>
</tr>
<tr>
<td>12.</td>
<td>Google (Play store, Android, Fuchsia)</td>
<td>Apps and software</td>
<td>31/10/2019; 05/12/2019</td>
</tr>
<tr>
<td>13.</td>
<td>Facebook</td>
<td>Social media</td>
<td>27/11/2019</td>
</tr>
<tr>
<td><strong>BUSINESS USERS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>Business user 2</td>
<td>E-commerce: Amazon</td>
<td>18/10/2019</td>
</tr>
<tr>
<td>16.</td>
<td>Business user 3</td>
<td>E-commerce: Amazon</td>
<td>20/11/2019</td>
</tr>
<tr>
<td>17.</td>
<td>Business user 4</td>
<td>E-commerce: Amazon</td>
<td>26/11/2019</td>
</tr>
<tr>
<td>20.</td>
<td>Business user 7</td>
<td>E-commerce: Etsy</td>
<td>04/12/2019</td>
</tr>
<tr>
<td>21.</td>
<td>Business user 8</td>
<td>E-commerce: eBay, Etsy</td>
<td>02/12/2019</td>
</tr>
<tr>
<td>23.</td>
<td>Business user 10</td>
<td>E-commerce: Undisclosed</td>
<td>18/12/2019</td>
</tr>
<tr>
<td>27.</td>
<td>Business user 14</td>
<td>Accommodation/ Hospitality: Booking.com</td>
<td>09/12/2019</td>
</tr>
<tr>
<td>28.</td>
<td>Business user 15</td>
<td>Accommodation/ Hospitality: Booking.com</td>
<td>06/12/2019</td>
</tr>
<tr>
<td>29.</td>
<td>Business user 16</td>
<td>Accommodation/ Hospitality: Booking.com</td>
<td>04/12/2019</td>
</tr>
<tr>
<td>30.</td>
<td>Business user 17</td>
<td>Accommodation/ Hospitality: Booking.com, Expedia</td>
<td>09/12/2019</td>
</tr>
<tr>
<td>32.</td>
<td>Business user 19</td>
<td>Apps and software: Google</td>
<td>17/10/2019</td>
</tr>
<tr>
<td>33.</td>
<td>Business user 20</td>
<td>Apps and software Undisclosed</td>
<td>20/11/2019</td>
</tr>
<tr>
<td></td>
<td>Business user</td>
<td>Apps and software</td>
<td>Date</td>
</tr>
<tr>
<td>---</td>
<td>---------------</td>
<td>---------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>34.</td>
<td>Business user 21</td>
<td>Apps and software: Google Play, App Store, Facebook</td>
<td>15/10/2019</td>
</tr>
<tr>
<td>35.</td>
<td>Business user 22</td>
<td>Apps and software: App Store, Google Play</td>
<td>04/12/2019</td>
</tr>
<tr>
<td>36.</td>
<td>Business user 23</td>
<td>Apps and software: Google Play</td>
<td>14/11/2019</td>
</tr>
<tr>
<td>39.</td>
<td>Business user 26</td>
<td>Apps and software: Android (Google)</td>
<td>22/10/2019</td>
</tr>
<tr>
<td>40.</td>
<td>Business user 27</td>
<td>Apps and software: App Store</td>
<td>10/10/2019</td>
</tr>
<tr>
<td>41.</td>
<td>Business user 28</td>
<td>Apps and software: Android (Google)</td>
<td>07/10/2019</td>
</tr>
<tr>
<td>42.</td>
<td>Business user 29</td>
<td>Apps and software: Android (Google)</td>
<td>25/09/2019</td>
</tr>
<tr>
<td>43.</td>
<td>Business user 30</td>
<td>Apps and software: App Store</td>
<td>11/10/2019</td>
</tr>
<tr>
<td>44.</td>
<td>Business user 31</td>
<td>Apps and software: App Store, Google Play</td>
<td>11/09/2019</td>
</tr>
<tr>
<td>45.</td>
<td>Business user 32</td>
<td>Apps and software: Google, App Store, Microsoft Store, Facebook</td>
<td>17/09/2019</td>
</tr>
<tr>
<td>46.</td>
<td>Business user 33</td>
<td>Apps and software: App Store</td>
<td>18/09/2019</td>
</tr>
</tbody>
</table>

**DATA BROKERS**

<table>
<thead>
<tr>
<th></th>
<th>Apps and software</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>47.</td>
<td>Algopix</td>
<td>E-commerce</td>
</tr>
<tr>
<td>48.</td>
<td>Jungle Scout</td>
<td>E-commerce</td>
</tr>
<tr>
<td>49.</td>
<td>AirDNA</td>
<td>Accommodation/ hospitality</td>
</tr>
<tr>
<td>50.</td>
<td>Apptopia</td>
<td>Apps and software</td>
</tr>
<tr>
<td>51.</td>
<td>SimilarWeb</td>
<td>Other</td>
</tr>
<tr>
<td>52.</td>
<td>Zirra</td>
<td>Other</td>
</tr>
</tbody>
</table>

**ASSOCIATIONS**

<table>
<thead>
<tr>
<th></th>
<th>Apps and software</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>53.</td>
<td>Computer &amp; Communications Industry Association (CCIA)</td>
<td>General: online platforms</td>
</tr>
<tr>
<td>54.</td>
<td>EDIMA</td>
<td>General: online platforms</td>
</tr>
<tr>
<td>55.</td>
<td>Digital Europe</td>
<td>General: online platforms</td>
</tr>
<tr>
<td>56.</td>
<td>The Marketplace Coalition</td>
<td>E-commerce: online platforms</td>
</tr>
<tr>
<td>57.</td>
<td>Centre for data innovation</td>
<td>General: online platforms</td>
</tr>
<tr>
<td>58.</td>
<td>Allied for Startups</td>
<td>General: online platforms</td>
</tr>
<tr>
<td>60.</td>
<td>Developers Alliance</td>
<td>Apps and software: business users</td>
</tr>
<tr>
<td>61.</td>
<td>European Games Developer Federation (EGDF)</td>
<td>Apps and software: business users</td>
</tr>
<tr>
<td>62.</td>
<td>Interactive Software Federation of Europe (ISFE)</td>
<td>Apps and software: business users</td>
</tr>
<tr>
<td>64.</td>
<td>EU Travel Tech (ETTSA)</td>
<td>Accommodation/ hospitality: business users</td>
</tr>
<tr>
<td>65.</td>
<td>Independent Retail Europe</td>
<td>E-commerce: business users</td>
</tr>
<tr>
<td>66.</td>
<td>The European Broadcasting Union (EBU)</td>
<td>Apps and software: business users</td>
</tr>
</tbody>
</table>
**Reference list**

**Academic**


### International organisations


Consultancies/think-tanks


Government agencies


Financial times. (n.d.). Amazon’s ever-increasing power unnerves vendors. Retrieved from https://www.ft.com/content/c82ce968-eb8a-11e8-94b2-17176fbf93f5


Mirani, L., & Nisen, M. (2014). The nine companies that know more about you than Google or Facebook. Quartz. Retrieved from https://qz.com/213900/the-nine-companies-that-know-more-about-you-than-google-or-facebook/


**Company websites (platforms, data brokers, others)**


Amazon Services Europe. What you should know about the Amazon MWS Reports API section. Retrieved from http://docs.developer.amazonservices.com/en_UK/reports/Reports_Overview.html


#:~:targetText=We%20gather%20it%20from%20a,their%20sales%20info%20with%20us.&targetText=This%20means %20that%20as%20we,analyze%20and%20test%20it%20daily.


