Polish Economic Institute

How much is our data worth?

JANUARY 2022

WARSAW

ISBN 978-83-66698-61-1

Citations:
Grzeszak, J., Łukasik, K., Święcicki, I. (2022), *How much is our data worth?*, Polish Economic Institute, Warsaw.

Warsaw, January 2022
Authors: Jacek Grzeszak, Krystian Łukasik, Ignacy Święcicki (PIE)
Cooperation: Agnieszka Wincewicz-Price, Paweł Śliwowski (PIE), Michał Bylicki, Ewa Zawojska (WNE UW)
Substantive editing: Andrzej Kubisiak
Editing: Annabelle Chapman
Graphic design: Anna Olczak
Graphic collaboration: Tomasz Gałązka, Joanna Cisek
Text and graphic composition: Sławomir Jarząbek
Polish Economic Institute
Al. Jerozolimskie 87
02-001 Warsaw, Poland

ISBN 978-83-66698-61-1

Table of contents

Key numbers

87%	of respondents claim that tech companies know too much about us.
84%	of Poles believe that tech companies' activity should be subject to greater control.
PLN 4.025 billion	value of Polish users' data for Google in 2020.
PLN 2.196 billion	value of data from Poland for Facebook in 2020.
81%	of Poles care about what tech companies are doing with their data.
PLN 17.07 per month	how much the average Pole is willing to pay so that Facebook does not have access to data aggregated on the platform and from other sources.
PLN 14.10 per month	how much we are willing to pay so that Google cannot access our data, including that concerning our activity on other portals.
69%	of respondents believe that no website or app should charge fees for access, and 43% do not want to pay online service providers for better privacy protection.

Key findings

In this report, we sought to measure the value of the data that Polish Internet users generate on digital platforms (social media and Internet search engines). This value was estimated from two sides: firstly, in terms of the revenue that Polish users' data generates for digital platforms (Facebook and Google) and, secondly, in terms of the value that the users themselves assign to the data and privacy online. In addition, we conducted a survey that aims to check Poles' knowledge of and views on digital platforms.

Above all, **respondents' replies indicate that the average user expects monetary compensation for the current situation, in which digital platforms have access to all our data and display personalised adverts.** This means that Poles consider the *status quo*, in which we pay for digital services using our data, undesirable.

According to PEI's study, the **average Polish user is inclined to pay PLN 17.07 per month to prevent Facebook from having access to data** aggregated on the platform and from other sources. In the case of **Google, Poles would be ready to pay PLN 14.10 per month to prevent it from accessing their data, including their activity on other Internet portals.**

At the same time, the platform's business model is based on processing this data and generating revenue from personalized adverts that are displayed to users. In the case of Facebook, ad revenue accounts for 98% of the company's revenue: USD 84 billion globally in 2020 (SEC, 2020a). At Google, this share is 80%: USD 146 billion globally in 2020 (SEC, 2020b).

According to PEI's calculations, for Google and Facebook, revenue from Polish users' data is **significantly higher than that reported by these companies' branches for the purpose of statistics and for the tax authorities.** For Google, monthly revenue from a single Polish user's data amounts to PLN 10.16. We therefore estimate that its **total revenue in 2020 from all its Polish users' data was PLN 4.025 billion.** For Facebook, monthly revenue from a Polish user's data amounts to PLN 8.52. **This means that – according to our calculations – total revenue in 2020 from all its Polish users' data was PLN 2.196 billion.**

In the survey, over half of Internet users (63%) agrees with calls for a ban on displaying ads based on private persons' data. This step would put an end to behavioural targeting by ads. This would block platforms' current business model, in which users pay for a service with their privacy. In addition, during the empirical part of the study, it turned out that **respondents value personalized ads over non-personalised ones and, in certain cases, would expect compensation for the lack of personalized ads.**

In addition, **most respondents are concerned about digital giants' growing influence – 84% believe that tech companies' activity should be subject to greater control.** Moreover, as many as 87% believe that digital platforms know too much about us. 77% of Poles are aware that they are paying for free services with their data. However, just 38% of respondents are ready to pay the providers they use for better privacy protection. This may be linked to their distrust towards these companies – 76% do not believe that a paid version of Facebook would better protect their rights. In the case of Google, this is 73%.

Introduction

This study by PEI is the first comprehensive effort to approach the problem of the value of data from two sides. On the one hand, we examined the value generated by the largest digital platforms – that is, Facebook and Google – from Polish users' data. We concentrated on these two companies because they are widely used by Polish Internet users and also constitute a point of reference for many other digital services. Almost 97% of Polish Internet users use Google (GlobalStats, 2021) and as many as 89% use Facebook (DataReportal, 2020). The two most-visited domains in Poland are "google.com" and "facebook.com" (Interaktywnie.com, 2020).

The two companies' revenue largely comes from using the raw material that is the data generated by users and processed by the platform. The writing of posts and rating, commenting on or searching for content by users providers the information the platform needs to sell its advertising products. The data generated by users is a guarded good and, with a few exceptions, cannot be exchanged on the market. This is why they do not have a clearly-defined price. Yet the lack of price does not mean that they have no value. One of the aims of our study is to estimate the value of the data that Polish users generate for global digital platforms.

On the other hand, we decided to juxtapose the value that users generate for the platform with a study of the value that the users assign to their own privacy, broadly understood. For this purpose, we carried out a discrete choice experiment, in which respondents were repeatedly presented with a set of choices between various versions of a service, from which they chose their preferred option. We studied users' inclination to pay for a service in a modified version – without platforms tracking their online activity, with a total lack of access and the inability for data left of the platform to be processed, without targeted advertising, and without creating a psychometric profile of the user. Juxtaposing these two perspectives enabled us to examine the distribution of benefits for the platform and its users.

The study was conducted in October 2021 in the form of a survey using the CAWI method on a sample of 944 people[1]. As part of the study, we also collected information on Poles' knowledge of the ways in which digital platforms work, their beliefs on given services' social utility and private usefulness, their opinions on regulation, their sense of control over their own data, and their readiness to pay for online services.

In the first chapter, we present the value of data in today's economy, especially in the context of the platform business model's sharp rise in the popularity. We include a review of the literature on the value of data. In the second chapter, we present the results of analysis of the value of Polish users' data for digital platforms. In chapter three, we present the methodology and result of our study of the subjective evaluation of data and privacy by Polish Internet users. The fourth chapter contains the results of the survey on Poles' knowledge and opinions on digital platforms. In the final chapter, we discuss the consequences of the PEI's study.

[1] Representative with regards to gender, age group and level of education.

The value of data in the contemporary economy

What is data?

The rapid development of ICT technology means that the subject of data occupies a central place in analysis of contemporary economies. Above all, advancing digitalization has meant that an ever growing number of human activities leave behind traces in the form of quantified information – data. Moreover, with technology progress and computers' exponentially-increasing computing power, the cost of storing, processing and transferring data has fallen dramatically. The resource's greater availability and decrease in the cost of processing it has contributed to the increased demand for data-based services.

↘ Box 1. Data as the new oil?

The growing role of IT resources means that analogies to traditional and familiar resources are often used to describe them. The most popular comparison is to oil.

- Like oil, data needs to be processed (refined) or turned into other raw materials. Raw data does not offer much value to its owner. Data also drives the contemporary economy, and controlling it is becoming almost as important as controlling access to hydrocarbons.
- Yet these resources have more differences than similarities: above all, data is irreplaceable. One set of data cannot be replaced by another, as it contains different information. Data is also a non-rivalrous good. This means that., unlike oil, consuming it does not reduce its quantity. Data can also be duplicated without losing its quality. With the digitisation of the economy and social life, data resources are constantly growing, unlike the limited and shrinking amount of natural resources.

To continue the energy analogy, data can also be compared to **renewable materials.** There is also an excess of these; **the challenge is to collect and process them appropriately and to match supply and demand** (in time and space). The key limitation of this comparison is above all the variety of data; although it is available in ever larger amounts, its irreplaceability means that we need to speak of growing diversity.

Most probably, there is no good analogy for data based on familiar resources. The ongoing legislative work in many countries seeks to define the rules for obtaining, processing and using data shows that new frameworks are needed to describe the new economy, corresponding to its unique needs.

Treating data as a factor of production, or one of the types of capital, we need to draw attention to a few of its properties:

- → irreplaceability – each set of data contains different information, and sets cannot be replaced without losing value. This differentiates them from traditional resources, such as energy commodities;
- → non-rivalry – a given set of data can be used multiple times and simultaneously by various entities, without losing their properties. Physical (machines, raw materials) or human resources can only be used in one place at once;
- → its value can only be assessed once it has been used – when deciding to invest in data (collecting and analysing it), there is no certainty about if and how it will provide benefits. In the case of traditional resources, the value stems from possessing them and their properties can be checked before acquiring them;
- → its price tends towards zero – this is because the platform can estimate the data of every successive user "of the same type" based on the data of users who have already made their data available (Acemoglu et al., 2019).

Data is highly differentiated, not only because it comes from different entities. In Table 1, we present the classification of data based on its source, owner and the type of access. Each of these types of data is subject to different regulations and can be used in different ways in the economy. In this report, we focus primarily on data generated by users and processed by the private sector.

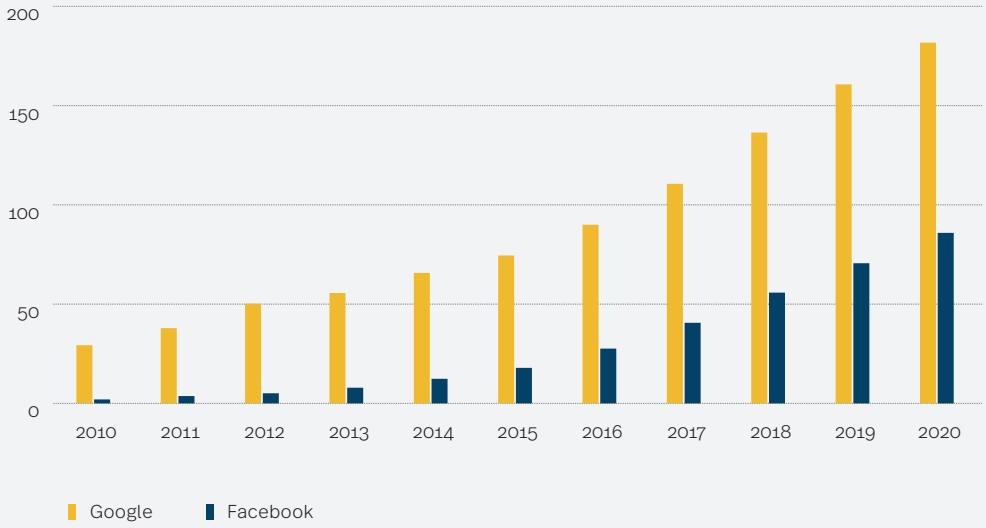
↘ **Table 1. Selected kinds of data, based on type**

Source	
Personal data	Information making it possible to identify an individual who is alive, such as: name and surname, email address, IP address, national identification number, police and medical records.
Organisational data	Information collected and processed by organisations (in both the public and private sector). It reflects a given institution's nature.
Owner	
Public sector data	Information collected, stored, created and processed by public institutions, such as open public data.
Private sector data	Information collected, stored, created and processed by private enterprises.
Type of access	
Proprietary data	Information protected by intellectual property law, such as patents, trade secrets and copyrights.
Public data	Information not protected by intellectual property law.

Source: prepared by PEI based on: Śledziewska, Włoch (2020).

Data has become a key factor of production in the contemporary economy, reflected in the ranking of the world's most valuable companies. Tech companies – including ones whose value is based on collecting, processing and monetizing data, like Alphabet (Google) and Meta (Facebook) – currently occupy the most places in the top ten companies with the highest market capitalisation (CompaniesMarketCap, 2021). Two or three decades ago, tech companies' key resource was still hardware, and the companies with the top market capitalisation included General Electric or Cisco. Today, hardware is available as a service, and advantage is conferred by intangible assets, which of course includes data. In 2015, intangible resources, including software and data, already accounted for 84% of the value of companies in the S&P 500 index (MIT Technology Review, 2016). The ability to analyse large datasets to optimise a company's internal processes, increase sales, plan the use of resources better or improve the quality of customer service determines a company's competitive position on the market (Śledziewska, Włoch, 2020). More importantly, as data has become a key factor of production, the platform business model – whose development is based on extracting datasets – has become more important. In 2020, the total revenue of Google and Facebook amounted to around USD 266 billion, which accounts for around 0.3% of global GDP.

↘ Chart 1. Annual revenue of Google (Alphabet) and Facebook (in billions of USD)



Source: prepared by PEI.

Digital platforms

The platforms' business model is based on acting as an intermediary between two separate but complementary groups of customers (Doligalski, 2013). The platform not only positions itself between the two sides of the market, but also constitutes the infrastructure needed for interaction between them to come about. In the cast of digital platforms, this kind of architecture ensures privileged access to the stream of data constantly generated by the platform's customers. The growing role of the platform model may be visible in how more non-tech companies are moving towards this model. For example, tractor producer John Deere has created a platform connecting producers of seeds, producers of chemical substances, farmers and hardware sensors. The data that the company acquires while acting as an intermediary between these groups is used to improve its products and services provided to clients, among other things (Srnicek, 2017). The move towards platforms results from the fact that they make better use of economies of scale, are "lean", create new sources of value and use data effectively to create positive feedback, among other things.

↘ **Table 2. Selected types of digital platforms**

Type of platform	Parties	How it operates
Advertising platforms (e.g. search engine, social network)	Users, advertisers	Obtaining platform users' data to sell advertising space. The cost is borne by advertisers purchasing personalised ads.
Service platforms (e.g. Uber, Airbnb)	Users, service providers	Matching users with service providers. Customers (who pay for the platform's operation) have more choice and service providers have access to a larger customer base and flexibility in managing resources.
E-commerce (e.g. Amazon, Allegro) and sales platforms (OLX, Vinted)	Buyers, sellers	Extending the existing market for exchange between sellers and buyers. They earn a percentage of the transaction commission, from the seller (e-commerce) or the buyer (certain sales).

Source: prepared by PEI.

The platform model's popularity stems from characteristics of an economy based on software and data, connected by the Internet. The latter means that more and more people and institutions are connected via a network that allows all sorts of communication. The use of software and data enables activity to be scaled up at almost no marginal cost. The platforms use network effects – both direct and indirect – and skilfully use data to engage users. By combining these effects, platforms can grow rapidly and monopolise (or oligopolise) the market. The winner (the monopolist or oligopolist) receives a sizeable reward.

↘ Box 2. Network effects

In the case of services, a network effect refers to when the user's benefit from a given service depends (positively) on the number of other users of the same service. In the case of digital platforms, network effects can be direct or indirect.

A direct network effect means that, as the number of users increases (in one of the two groups served by the platform), the utility of users in the same group increases. The best examples are a telephone network or social media platform – with every successive user, a given platform's utility for the other users increases, as they are able to form a larger number of connections.

An indirect network effect occurs when the increase in the number of users of one of the groups on the platform increases the utility of another group. For example, as the number of users of a platform offering software increases, the utility for developers putting their solutions on it increases. In contrast, the impact of each additional user on the programmes' users is minimal or zero. In other cases, it can even be negative – on an auction platform, each additional seller increases the benefit for buyers (more choice), but can lower the benefit for other sellers (more competition).

Among the top ten companies in terms of market capitalisation, as many as four [Alphabet (Google), Amazon, Meta (Facebook), Tencent] are companies that largely owe their position to the skilful use of the platform model and of data. The change in paradigm when it comes to how the private sector generates value may also be visible in the number of terms that seek to make sense of the changes taking place in the digital economy. Terms such as "the gig economy", "the sharing economy", "the attention economy" or "surveillance capitalism" all draw attention to various aspects of the domination of the platform model of organising a business.

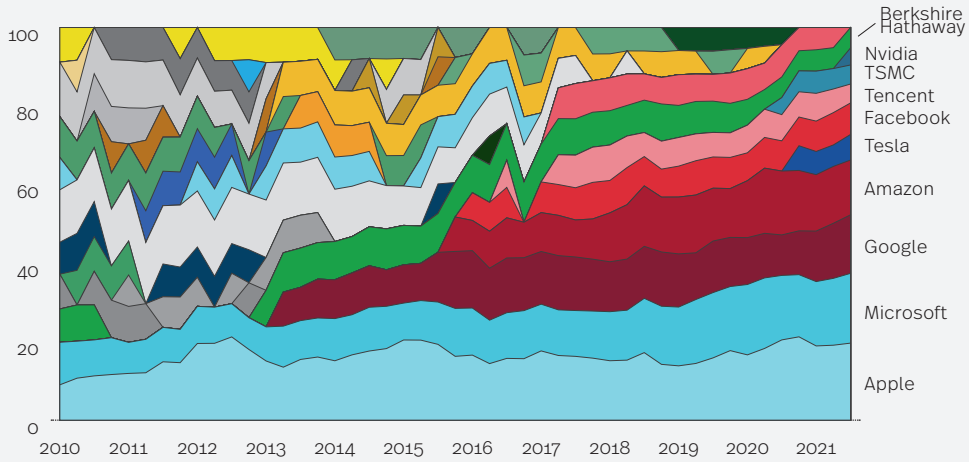
Chart 2 shows the sharp increase in platforms' importance. In Q3 2021, platforms had a 42% share in the top 10 publicly listed companies with a highest market capitalization in the world. For comparison, ten years ago, this share was zero. Digital platforms are marked in red, tech companies in blue, financial institutions in green, the energy and mining industry in grey, and production and FMCG companies in yellow.

Another feature characterising platforms are the sources of financing, which are constructed differently than on the traditional market. Some platforms only charge one of the parties; the one with less price elasticity or that is more dependent on the availability of users on the other side (such as Facebook, Google and the free version of Spotify). The service is therefore subsidised for some of its users. Platforms also benefit from unique knowledge about demand for products and can modify prices in real time – like Uber, whose drivers have no influence over the price of a journey and, by implication, their revenue.

Here, it is worth noting that the lack of financial charges does not mean that the users do not bear any expense for using the service offered by the platform. That expense is the data that they transfer directly to the platform (such as posts on a social media platform, or the choice of link in their search results) or information that the platforms obtain without the users having to do anything (such as their location or the model of their device). This data is then used

to create a profile of the user, often using statistical techniques that enable him or her to be assigned characteristics that he or she would not want to reveal (for example, sensitive data concerning his or her sexual orientation, political views or health).

↘ Chart 2. **Cumulative values of the top 10 publicly listed companies in the world by stock market capitalisation (as a percentage)**



RED - Google, Amazon, Facebook, Alphabet, Alibaba Group, Tencent
BLUE - China Mobile, Tesla, IBM, Nvidia, TSMC, Samsung Electronics, Microsoft, General Electric, Apple
YELLOW - Procter & Gamble, Johnson & Johnson, Walmart, Novartis, Hoffmann-La Roche, Nestlé
GREEN - JP Morgan Chase, Wells Fargo, ICBC, China Construction Bank, Berkshire Hathaway, Visa, AT&T
GRAY - Exxon Mobile, PetroChina, Petrobras, Chevron Corporation, BHP Billiton, Royal Dutch Shell

Source: prepared by PEI.

Access to this kind of information (including intimate information) enables the platform to create a psychometric profile of the user and adapt advertising to him or her – not only based on demographic data or location, but also based on a given person's views, values and fears. Behavioural products (understood as ads or other messages that use information about the user) based on sensitive data about the users were also used during significant political events, such as the Brexit referendum or elections in countries around the world (Cadwalladr, 2020).

For the purposes of this report, we will examine how platforms in which the service is free for users are financed from advertising revenue. The Diagram 1 illustrates this mechanism: the flow of data and financial resources on the platforms.

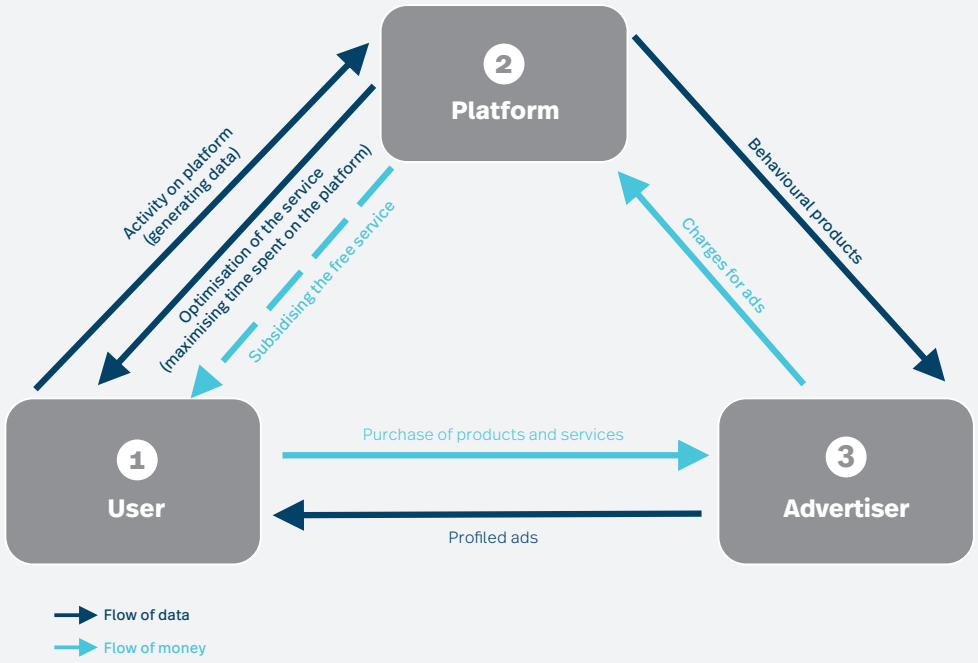
Diagram 1 shows the flow of data and money on advertising platforms:

1. By sharing photos, reviewing restaurants, liking posts or using an online search engine, the user generates data for the platform. Each time he or she is active online,

he or she produces additional raw unstructured information that results from any kind of actions online (such as a series of clicks or moving the cursor). The service on the platform is seemingly free for the user; that is, he or she does not cover the cost of the service with money. In this sense, the platform subsidises the service on the user's side.

2. On the one hand, this data, which Zuboff (2020) calls behavioural data, is used by the platform to improve the quality of the service it is providing (such as the accuracy of Google search results). On the other hand, the platform uses it to sell advertisers behavioural products.
3. The advertisers bear the financial cost of displaying an ad, which depends on the scope of the data used to define the target group. This cost is in principle the platform's only source of income. These kinds of products allow advertisers to reach groups of users that fit very narrow criteria with their marketing message. When users buy the products being advertised, the advertisers make money.

↘ Diagram 1. Flow of data in the platform economy



Source: prepared by PEI.

Review of the literature

In the literature, the analysis of the value of data concentrates on attempts to estimate the impact of free digital services on the size of GDP (Bukht, Heeks, 2018; IMF, 2018; Nakamura, Samuels, Soloveichik, 2016) or to quantify the impact of free digital services on social welfare (Brynjolfsson, Collis, Eggers, 2019; Brynjolfsl et al., 2018; Bughin, Manyika, 2013). Another popular approach is estimating the value that users assign to their data or privacy. These goods cannot be exchanged on the market, so survey and empirical methods are used to estimate their value. According to this approach, the average American is inclined to pay USD 5 per month to protect his data and would want to receive USD 80 for access to this day (Winegar, Sunstein, 2019). Overall, for the use of different types of data, users would demand from USD 1.82 (access to their location) to USD 8.44 (access to data from their bank account) per month (Prince, Wallsten, 2020). Maciej Sobolewski and Michał Paliński (2017) estimate that the value resulting from the increase in the level of privacy thanks to the GDPR amounts to EUR 6.5 per month.

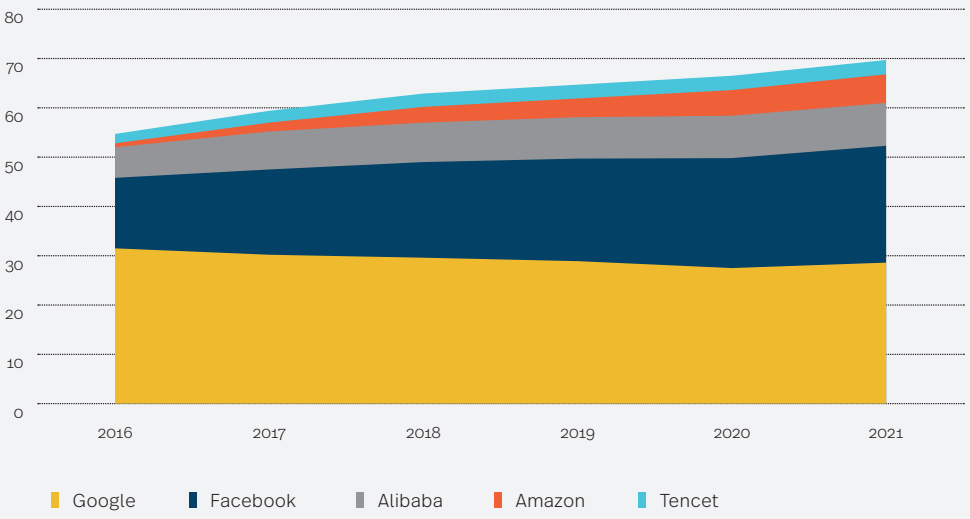
From companies' perspective, the analysis of the value of data focuses on calculating the value that stems from open data (Bennett Institute, 2020), the higher market value of companies that invest in data (PWC, 2019) or the income that comes from using behavioural data to target advertising (Acquisti, Marotta, Abhishek, 2019). On the business size, Shapiro and Aneja (2019) and Facebook (SEC, 2020a) have used an approach similar to ours. Yet unlike the above, this analysis takes into account different amounts of revenue, depending on geographical region, and does not generalise by applying the American context to other regions.

The value of data for digital platforms

In our study, when estimating the value that the data of a single user in Poland generates for online search engines and social media platforms, we concentrated on Google and Facebook. As noted already, Google and Facebook reach the vast majority of Polish Internet users: almost 97% of them use the Google search engine (GlobalStats, 2021) and 89% use the Facebook social media platform (DataReportal, 2020). Moreover, globally, these two companies together have an almost 50% share of the online advertising market, which is their main source of income (eMarketer, 2021) (Chart 3).

The business model of Facebook social platform and Google services analysed by us is based on generating revenue from ads displayed to users. In the case of Facebook, advertising revenue constitutes 98% of its revenue – USD 84 billion globally in 2020 (SEC, 2020a). In the case of Google, this share is 80% – around USD 147 billion globally in 2020 (SEC, 2020b).

↘ Chart 3. **Selected companies' (Internet platforms') share in global online advertising revenue (as a percentage)**



Source: prepared by PEI based on: Marketer (2021).

Analysis of the value of data for digital platforms

In this analysis, we concentrate on calculating the average monthly revenue generated by a Polish user for a given platform. After considering the approaches available, we chose the method used both in independent studies (Shapiro, Aneja, 2019) and in analyses conducted by digital platforms (SEC, 2020a).

We estimated the average monthly revenue generated for Google and Facebook by a single user's data in the same way.

In both cases, we obtained the data on advertising revenue from the annual financial reports summitted to the US Securities and Exchange Commission (SEC). This enabled us to separate revenue from users' data (in our study, equated with revenue from advertising) from other sources. For Facebook, we were able to use data on its revenue from Europe[2] (SEC, 2020a). For Google, we used global data (SEC, 2020b).

Next, for Google, this amount was divided by the global number of Internet users (World Bank, 2020a) to estimate the average revenue from a single Internet user's data. Google does not reveal how many people use its services; we used the number of Internet users as a good proxy. In the third step, we adjusted the revenue from a single user for differences in individual countries' wealth. For this, we used data on the GDP per Internet user in Poland compared to the global average (World Bank, 2020b). On a global and Polish scale, the proportion of Internet users who use Google is similar, hence we equated the number of Google users with the number of Internet users in general.

In the case of Facebook, there are significant differences in the percentage of Internet users who use the platform. Here, we used data on the average revenue per user (ARPU) for users in Europe and the total revenue from this region to estimate the number of Facebook users in Europe (the company does not directly report these numbers) (SEC, 2020a). Then we divided revenue from the region by the number of users and, as in the case of Google, adjusted it for relative differences in wealth (measured in terms of GDP per Facebook user in a given region).

Based on these assumptions, we calculated that:

- → **For Google, average monthly revenue from a single Polish user's data is PLN 10.16. Total revenue from all its Polish users' data amounted to PLN 4.025 billion in 2020.**
- → **For Facebook, average monthly revenue from a single Polish user's data is PLN 8.52. Total revenue from all its Polish users' data amounted to PLN 2.196 billion in 2020.**

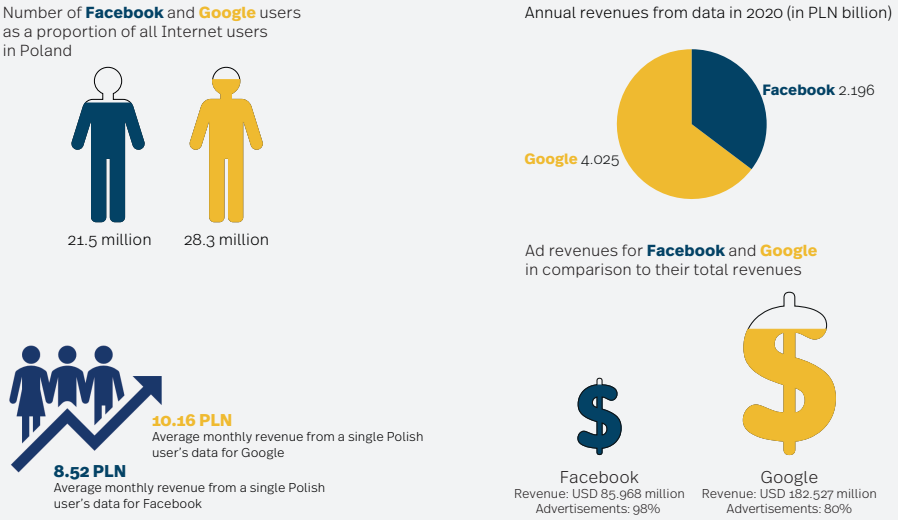
Facebook itself used a similar method to estimate the annual monthly revenue from a user's data. It reports that average monthly revenue per user in Europe is USD 4.25 per month (PLN 8.69)[3] (SEC, 2020a). However, this number takes into account all kinds of revenue, not just that from advertising activity.

Chart 4 shows the increase in average annual revenue from a user's data globally. The method used to calculate the revenue in a given year is analogous to the one used in our report. For Google, the compound annual growth rate (CAGR) over the course of the whole decade is 8.16%; for Facebook, it is as high as 25.23%.

[2] In its 10-K report, Facebook includes Turkey and Russia in its Europe region; this was taken into account in all the calculations.

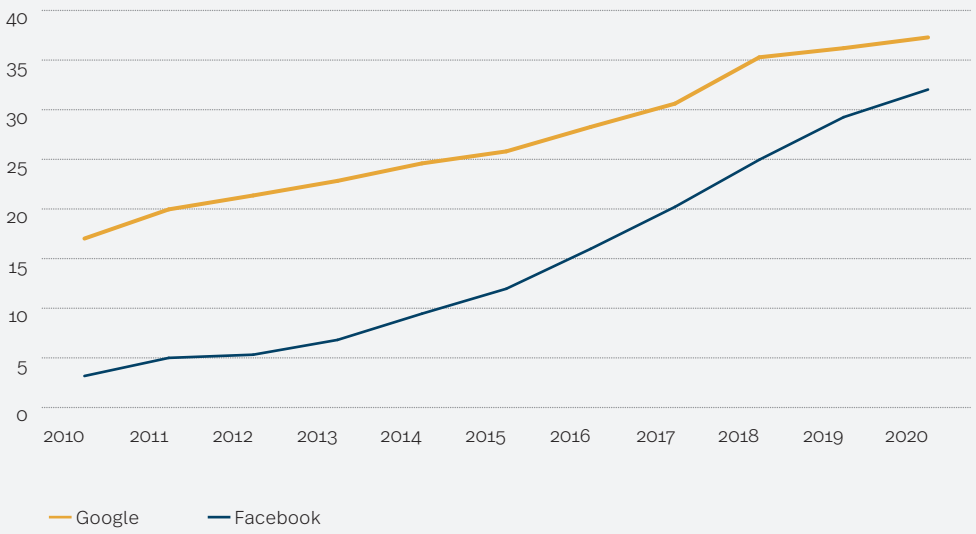
[3] Averaged amount from Q1-Q4 2020, adjusted for GDP per Polish Internet

↘ Infographic 1. Facebook and Google's revenues from Polish users' data in 2020



Source: prepared by PEI.

↘ Chart 4. Average annual revenue from the user's data (globally, in USD)



Source: prepared by PEI.

What value do Poles assign to their data?

Methodology

Unlike market goods, the price of which can be seen in commercial transactions, user preferences for the protection of their privacy remain hidden. The purpose of the study conducted for the purpose of this report was to measure how users value the hypothetical privacy protection on Facebook and in Google's services. To this end, we used non-market stated preferences survey in our study. We used data obtained from specially-designed surveys that contain hypothetical choice situations in the form of Discrete Choice Experiment (DCE). The questions in the survey reveal the value that a given person attributes to goods (or services), where the good is characterised by separately-valued attributes, and each respondent is asked several times to choose the best option available, taking into account its features and cost. The second part of the study was a survey on Poles' beliefs about digital services. This was supplemented by questions testing respondents' knowledge of how digital services work and demographic data.

↘ **Diagram 2.** Example of a set of DCE options used in our study (version for Facebook)

Variant	A	B	C
Privacy	Has access to the data placed on the platform	Does not have access and does not analyse any of your data	Has access to data, also from activity beyond Facebook
Profiling	Does not profile you	Does not profile you	Profiles you
Ads	Ads	No ads	Ads based on your data
Cost	5 PLN per month	20 PLN per month	0 PLN per month

Source: prepared by PEI.

The presented variants of choice concerned four attributes: privacy, described as the service provider's access to data (both that placed directly on the platform and that obtained from other websites visited); profiling (determining the user's hidden features based on statistical analysis and his or her activity on the platform); the presence of advertisements (in general and targeted ones), as well as the monthly cost of a given variant.

When assessing the value of non-market goods using models based on declared preferences, one can test users' willingness to pay (WTP) or willingness to accept (WTA) for access to or restriction of access to a given good. In our study, WTP was selected partly due to the inflated results generated by the WTA method, which are caused by the "endowment effect", people's tendency to assign higher value to goods that they already have (Thaler, 1980). The method selected (DCE) allowed us to reduce the measurement inaccuracies in the declared preferences (Prince, Wallsten, 2020), as well as to extract values for individual levels of selected attributes. For better results, further, in-depth empirical research should be carried out. Introducing an actualising stimulus (for example, money) into the study would probably reduce the estimates presented. In other words, when respondents have to bear the real cost of protecting their data, the amount they are willing to pay is usually lower than that in research based solely on declarations. Nevertheless, the results of studies comparing survey methods with empirical ones prove that the values obtained in both cases correspond (Bizon, 2016).

↘ Box 3. Comparison of WTP and WTA

WTA (*willingness to accept*) – the limit sales price, the minimum monetary amount of compensation that a person is willing to accept in exchange for giving up a certain good.

WTP (*willingness to pay*) – the limit purchase price, a given person's inclination to buy a certain good for a certain amount.

A possible way to estimate the value of a non-market good in one of the two above approaches is a survey in which respondents are presented with a number of service variants with variation in the values of individual parameters.

In our experiment, a total of 18 variants were tested for Facebook and Google. Each participant was shown six sets of three variants each for Facebook and same number for Google. In each case, there was also a choice of the *status quo* variant, in which we assumed that the platform (Facebook or Google) obtains data both from the service offered and from other sources, creates a user profile, shows him or her profiled ads, and the entire service is free.

The DCE survey was conducted in computer assisted web interview (CAWI) format around the end of the third quarter of 2021 on a representative group of 944 participants. Extensive qualitative preliminary studies (in-depth interviews, pilot study) helped refine the

attribute descriptions and optimise the options available. Econometric preference analysis is based on random utility theory (McFadden, 1974). A random parameters (mixed) logit model (MLX), assuming a variety of preferences among respondents, was used to analyse the data. A similar methodology is commonly used in research on the valuation of non-market goods (Budziński, 2015; Paliński, 2021; Carson, Czajkowski, 2014; Train, Weeks, 2005; Sobolewski, Paliński, 2017). For detailed results of the model, see Appendix 1.

Results

Our study shows that **the average Pole is inclined to pay PLN 17.07 per month for Facebook not to have access to data aggregated on the platform and from other sources. Poles would be ready to pay PLN 14.10 per month for Google not to have access to private data, including activity on other websites.** These amounts can be interpreted as the value that the average Polish Internet user assigns to access to his or her data. For partial privacy – that is, only allowing the websites to access data on the platform or in users' search history, without access to data on their activity on other websites–respondents were willing to pay PLN 12.35 for Facebook and PLN 6.71 for Google. To avoid profiling by the platform – that is, the process of discovering features that the user had not previously shared on the platform (for example, about their sexual orientation) based on their online activity – users were ready to pay PLN 3.60 per month in the case of Facebook and PLN 1.92 in the case of Google. To avoid advertising on Facebook, they would be willing to pay PLN 3.81 per month; for Google, this is PLN 4.34. Interestingly, **the results of the survey show that users attribute positive utility to targeted ads. If they were to receive a non-personalised advertisement on the platform, they would expect compensation of PLN 1.28 from Google and PLN 1.04 from Facebook** (although, in the latter case, the result was statistically insignificant).

↘ Table 3. Willingness to pay a monthly charge for a given attribute (in PLN)

Attribute	Facebook	Google
Full privacy	17.07	14.10
Partial privacy	12.35	6.71
Lack of ads	3.81	4.34
Lack of targeted ads	-1.04*	-1.28
Lack of profiling	3.60	1.92

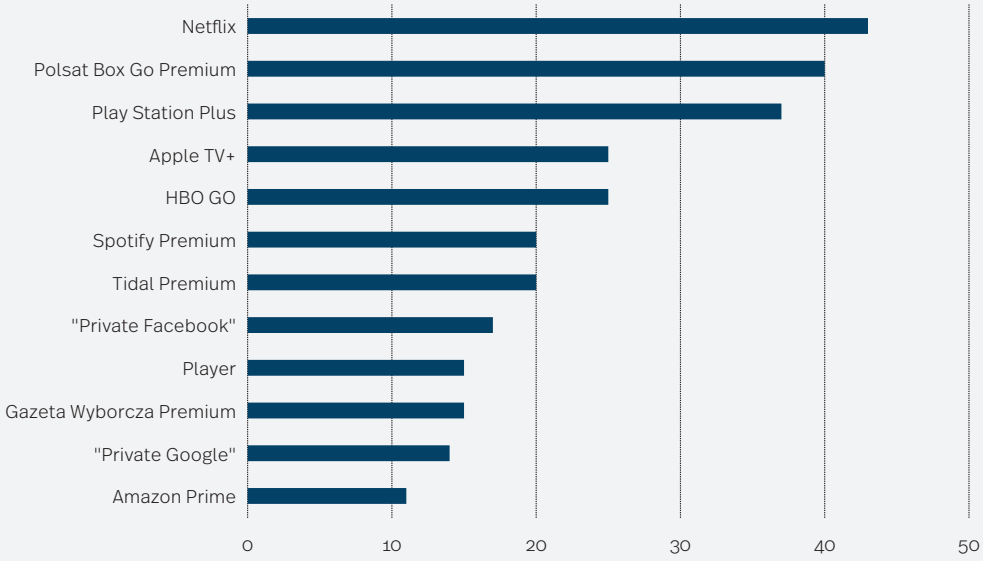
Note: * statically insignificant result.
Source: prepared by PEI.

Significantly, respondents' replies point to the negative value of the *status quo*. In other words, users expect compensation for the current situation, in which a platform has access to all our data, creates a profile of its users and displays personalised ads.[4]

In this way, the results of our study show that Polish Internet users would be willing to pay a monthly charge for a service in the form of a search engine or social network (on a similar basis to the monthly charge for a streaming service) if, in exchange, the platforms did not collect information about their users. **This opens the way to a discussion on a change in the model of how digital platforms function.**

In this context, it is also worth drawing attention to the charges incurred by users of digital services in Poland (Chart 5). More and more services are available based on a subscription model, in which users pay for access and ads are not displayed while they are using the service. Examples include VOD platforms, music streaming services (though a "free" version financed from advertising is often available, too), and – increasingly often – news services (Grzeszak, 2021).

↘ Chart 5. Monthly charge for selected digital services in Poland (in PLN)



Source: prepared by PEI.

[4] The *status quo* variant was defined in this way. Users currently have the ability to change their privacy settings, but the vast majority maintain settings that allow platforms to access a wide range of their data.

Comparison of the results

The value of Polish users' data for Google and Facebook estimated based on these two companies' revenue is significantly lower (in case of Facebook even more than half the value) that the survey respondents would be willing to pay for maximum privacy protection.

↘ **Table 4. Results of two types of analyses: revenue and DCE**

	Value of Polish users' data			
	For the companies		For Internet users	
Facebook	Monthly value of the average Polish user for Facebook	**PLN 8.52**	**PLN 17.07**	Value of the data made available to Facebook by the average user
Google	Monthly value of the average Polish user for Google	**PLN 10.16**	**PLN 14.10**	Value of the data made available to Google by the average user

Source: prepared by PEI.

In addition, this is a situation in which platforms generate significant revenue from users' data, while at least some of these users obtains negative utility from the current settings concerning the use of their data.

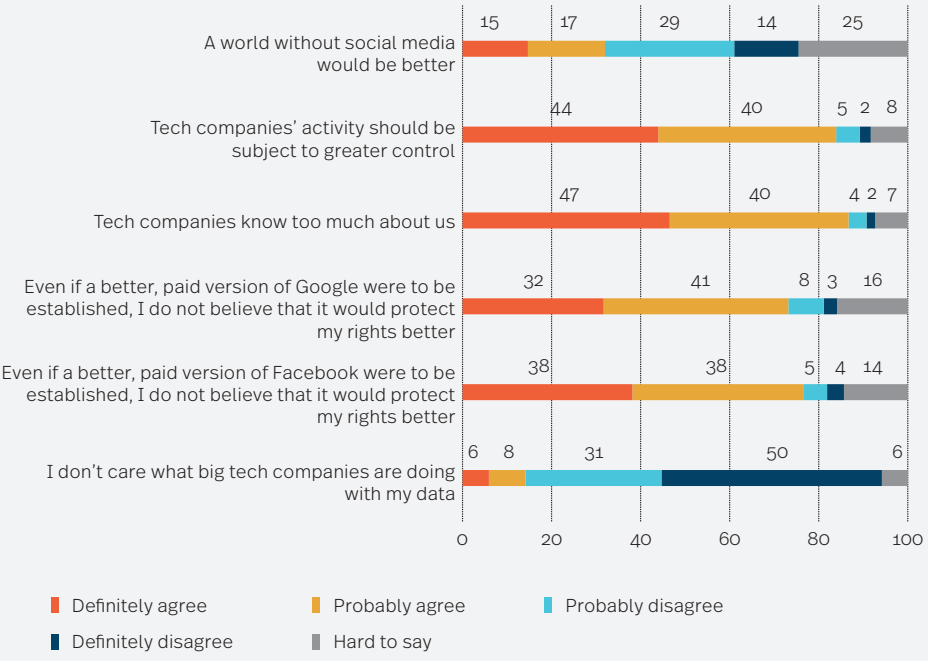
Comparing the results of these two analyses opens the way to a discussion on changes in the model that digital platforms use to function, which is outlined in the final part of this report.

Poles' views on digital platforms and services

Overview of the responses

The Polish Economic Institute's survey that accompanied the study on the value that users assign to their data concerned their views on digital services, online ads, and paying for online content and services.

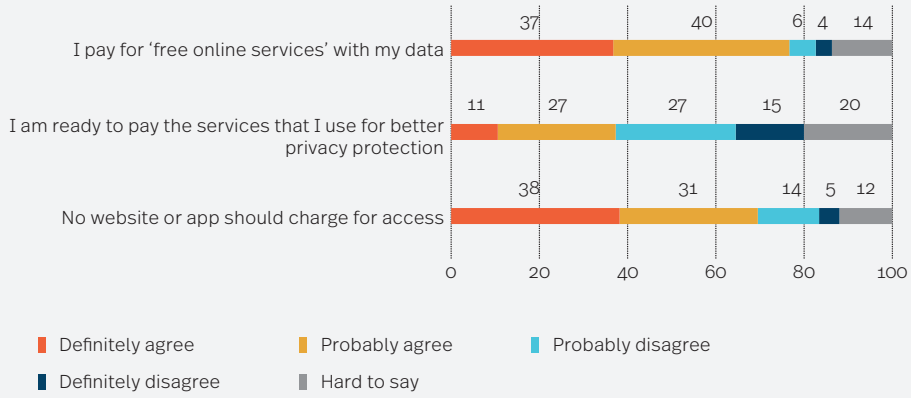
↘ Chart 6. **Respondents' attitude to selected statements concerning digital services and the companies that provide them (as a percentage)**



Source: prepared by PEI.

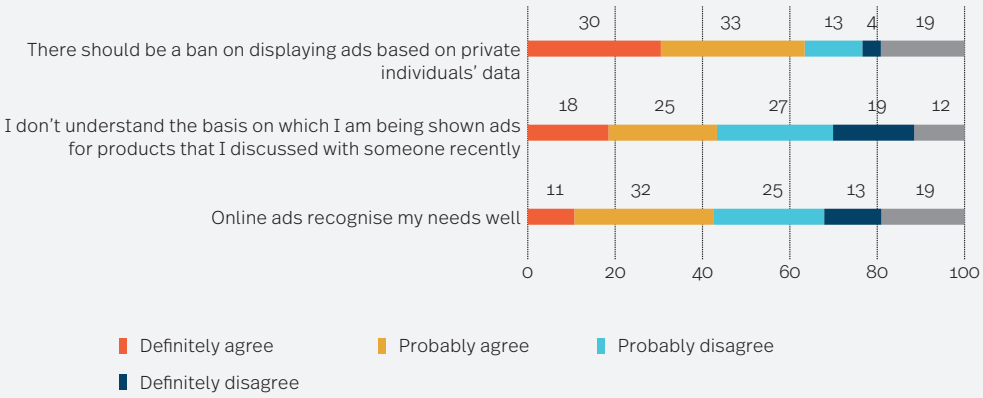
Most respondents are concerned about digital giants' growing influence: 84% believe that tech companies' activity should be subject to greater control, and 87% that these companies know too much about us. This may be related to the distrust towards specific companies: 76% of respondents does not believe that a paid version of Facebook would better protect their rights (73% in the case of Google). Most of respondents (81%) are not indifferent to what is happening to their data. It should be noted that this study was conducted before former Facebook employee Frances Haugen drew attention to Facebook's approach to problems generated by social media.

↘ Chart 7. **Respondents' attitude to selected statements concerning paying for digital services and online content (as a percentage)**



Source: prepared by PEI.

↘ Chart 8. **Respondents' attitude to selected statements concerning online ads (as a percentage)**



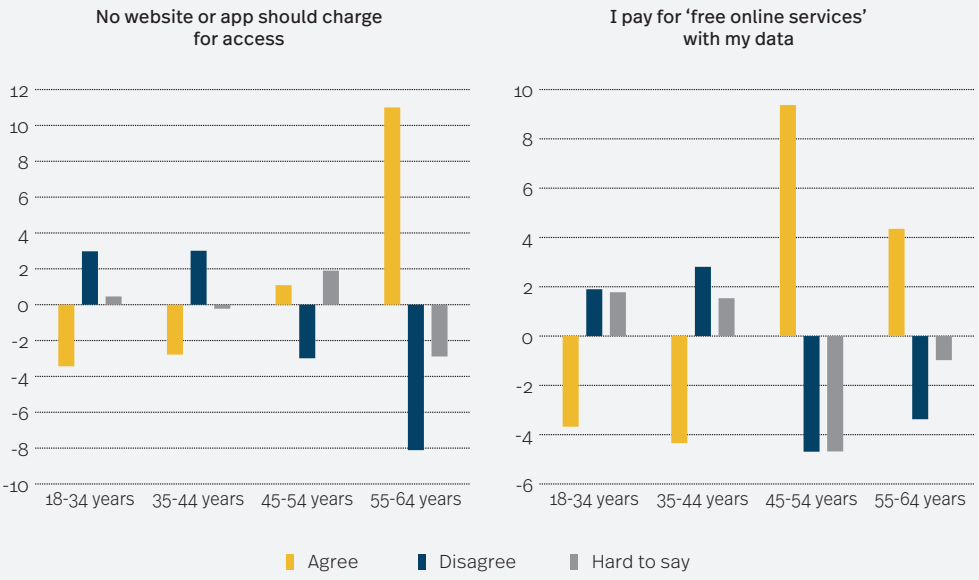
Source: prepared by PEI.

Most Internet users know about the transaction involving exchanging data for access to digital services described in this report. 77% of respondents agree with the statement that they actually pay for free online services with their data. 10% disagree and, for 14%, it is hard to say. This state of affairs seems to be widely accepted. 69% of respondents believe that no website or app should charge for access, and 43% do not want to pay websites for better privacy protection.

The responses to statements concerning online ads may seem paradoxical. 63% of Internet uses agree that there should be a ban on displaying adds based on private individuals' data. This ban would put a stop to the behavioural targeting of ads. As a result, the only way to adjust ads to users' needs would be to use that location from which someone goes online, or contextual advertising. However, 43% of respondents say that the current ads addressed at them respond to their needs, and 19% struggled to answer this question. 46 percent of respondents understand the basis on which they are shown ads for products.

Awareness of the transaction – data for access to services

↘ Chart 9. Respondents' attitude to selected statements, by age group (differences with regards to average for the whole population, in percentage points)



Source: prepared by PEI.

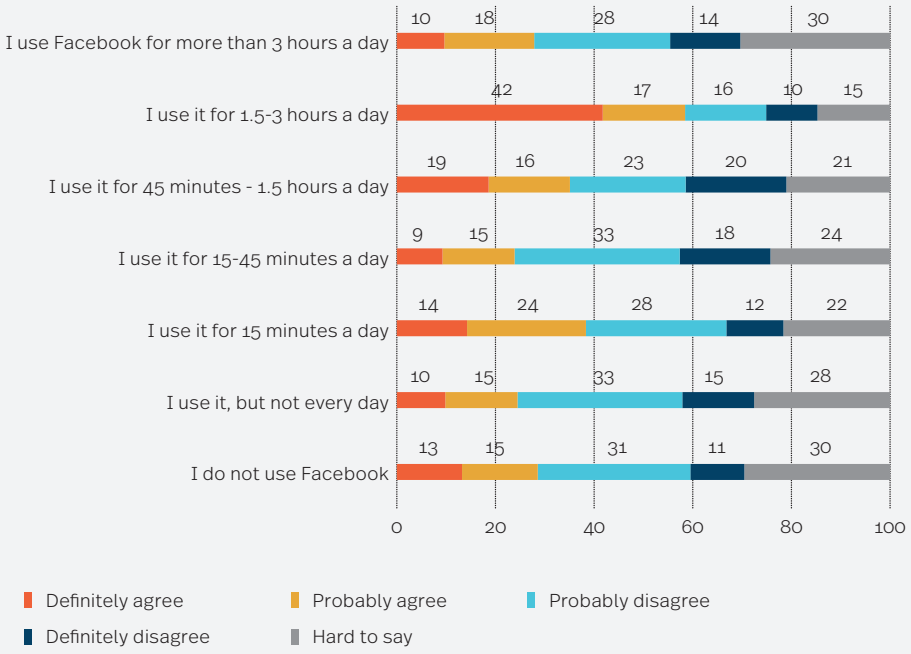
Older people are more convinced that the Internet is free: 81% of respondents in the 55-64 age group, 11 pp more than in the population as a whole. At the same time, however, older people were more likely to say that they understand the mechanism via which we pay for access to online services using our data. It is worth emphasising here that people in the younger age groups use paid online services (such as VOD, music services, and so on) more often (Grzeszak, 2021), which is in line with the lower percentage of respondents who expect the Internet to be entirely free.

↘ Chart 10. **Respondents' attitude to the statement "I pay for 'free online services' with my data", by level of education and earnings (as a percentage)**

Source: prepared by PEI.

People with secondary or university education, as well as people in the upper class based on revenue (over PLN 4500 net per household member), are more likely to be aware that we pay for digital services with our data.

↘ Chart 11. **Respondents' attitude to the claim "A world without social media would be better", by declared time per day using Facebook (as a percentage)**



Source: prepared by PEI.

Interestingly, people who say they spend 1.5-3 hours per day on Facebook tend to have a more critical attitude towards social media. In this group, 42% of respondents said that they strongly agree with the statement that a world without social media would be better. However, it is worth emphasising that respondents reported the time spent using the websites from memory and that it is possible that people who use social media a lot, while having a negative attitude towards it, systematically underestimated the perceived time they spent on the platform (this would explain the big difference in responses between people who use it for more than 3 hours a day and those who use it between 1.5 and 3 hours a day).

Discussion

Our study draws attention to three key issues that should become the foundation for a discussion on how digital platforms operate in Poland.

Firstly, according to this study, **the value of the data generated by Polish users is significantly higher than the amount reported by the companies Facebook Poland and Google Poland in their financial reports in the National Court Register (KRS).** In its financial report for 2020, Facebook Poland recorded PLN 724.14 million in revenue in 2020 and paid PLN 5.2 million in income tax. Meanwhile, in accordance with our calculations, the value of Polish users' data for Facebook amounted to PLN 2.2 billion in 2020. In its financial report, Google Poland[5] recorded PLN 546.52 million in revenue and paid PLN 12.8 million in income tax. Our calculations show that the value of Polish users' data for Google amounted to PLN 4 billion in 2020. We are therefore speaking of around a threefold and sevenfold difference, respectively. It is worth noticing here that the values calculated by us cannot, at the moment, be equated with the companies' revenue for the purposes of determining the amount of tax due. The difference between the declared revenue and the revenue arrived at in our study results from the fact that the Polish Economic Institute calculated the value that Polish users' data generates for each of the companies. In contrast, revenue from advertisers may come from all over the world, not only from Poland.

Secondly, Poles are dissatisfied with the *status quo*, in which they pay for digital services with their data. They are willing to pay PLN 17.07 per month for a Facebook service that would protect their privacy and PLN 14.10 for the analogous service from Google. At the same time, both companies generate on average PLN 8.52 and PLN 10.16 per month from a single Polish user's data. This means that, theoretically, **an alternative model for managing platform services could be created,** with Internet users paying a monthly subscription for a service that protects their privacy and does not display ads. For example, for a fee of around PLN 10 per month, both sides would feel the benefits; this amount is lower than that declared by respondents, but higher than the platforms' average monthly revenue from a single user.

Thirdly, Polish **Internet users' knowledge of how the platform economy works is surprisingly large.** As many as 77% of respondents know that they are actually paying for free online services with their data. At the same time, as the results of the DCE study show, they are dissatisfied with this state of affairs. This leads us to the **paradox visible in the results of the study:** asked directly, most respondents think that websites should be free and would be unwilling to pay to protect their privacy. Juxtaposed with respondents' lack of conviction that paid versions of the services would better protect their privacy and sense of threat caused by online surveillance, **the transformation of the platform model may require more radical change than introducing a subscription model.**

The discussion on this matter remains open and the subscription model mentioned earlier is not the optimal solution. Assigning

[5] The companies Google Cloud and Google Partners operate in Poland, too, but were not taken into account in this study.

ownership to the data that users generate can be problematic, as it is often difficult to identify clearly whom the information generated should belong to. Moreover, the low bargaining power of a single user compared to a global corporation would mean that data could be sold at overly low prices. Privacy can also be thought of as an inalienable right that should not be subject to market operations. In addition, in the subscription model, users would pay to access content that they themselves produce. This solution differs from that currently used by streaming services, where the fee serves to cover the costs of film production or paying musicians who present their content. In contrast – apart from the cost of maintaining and developing the service – Facebook does not produce its own content; it is attractive because of the content created by its users. Moreover, not all Poles could afford to subscribe to digital services that are currently free. In this sense, the advertising-based model is more democratic: it provides the same service to both richer and poorer users. For these reasons, further research, based on the Polish Economic Institute's findings, should consider possible models for maintaining digital services in a way that does not violate Internet users' right to privacy.

Bibliography

Acemoglu, D., Makhdoumi, A., Malekian, A., Ozdaglar, A. (2019), *Too Much Data: Prices and Inefficiencies in Data Markets*, NBER Working Paper Series, National Bureau of Economic Research, Cambridge.

Acquisti, A., Marotta, V., Abhishek, V. (2019), *Online Tracking and Publishers' Revenues: An Empirical Analysis*, Workshop on the Economics of Information Security, Boston.

Bennett Institute (2020), *The Value of Data*, Cambridge.

Bizon, W. (2016), *Ankieta a eksperyment w kontekście WTA i WTP – rola bodźca urealniającego w procesie wyceny danych osobowych*, "E-mentor", nr 5(67).

Brynjolfson, E., Collis, A., Diewert, W., Eggers, F., Fox, K. (2018), *The Digital Economy, GDP and Consumer Welfare: Theory and Evidence*, The Sixth IMF Statistical Forum, http://research.economics.unsw.edu.au/kfox/assets/session-4_kevin-fox_bcdef__digitaleconomy_imf_nov2018_presentation.pdf [accessed: 29.10.2021].

Brynjolfsson, E., Collis, A., Eggers, F. (2019), *Using massive online choice experiments to measure changes in well-being*, Proceedings of the National Academy of Sciences, No. 116(15), https://doi.org/10.1073/pnas.1815663116.

Budziński, W. (2015), *Znaczenie zmienności krańcowej użyteczności kosztu ponoszonego przez konsumenta dla wyceny dóbr nierynkowych*, "Ekonomia", nr 43.

Bughin, J., Manyika, J. (2013), *The mobile Internet's consumer dividend*, McKinesy, https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-mobile-internets-consumer-dividend [accessed: 29.10.2021].

Bukht, R., Heeks, R. (2018), *Defining, Conceptualising and Measuring the Digital Economy*, "International Organisations Research Journal", No. 13, https://www.researchgate.net/publication/327356904_Defining_Conceptualising_and_Measuring_the_Digital_Economy [accessed: 29.10.2021].

Cadwalladr, C. (2020), *Fresh Cambridge Analytica leak 'shows global manipulation is out of control'*, The Guardian, https://www.theguardian.com/uk-news/2020/jan/04/cambridge-analytica-data-leak-global-election-manipulation [accessed: 29.10.2021].

Carrière-Swallow, Y., Haksar, V. (2019), *The Economics and Implications of Data*, International Monetary Fund, Washington.

Carson, R., Czajkowski, M. (2014), *The discrete choice experiment approach to environmental contingent valuation*, (in:) Hess, S., Daly, A. (ed.), *Handbook of Choice Modelling*, Edward Elgar Publishing, Cheltenham.

CompaniesMarketCap (2021), *Largest Companies by Market Cap*, https://companiesmarketcap.com/ [accessed: 29.10.2021].

DataReportal (2020), *Digital 2020: Poland*, https://datareportal.com/reports/digital-2020-poland [accessed: 12.10.2021].

Doligalski, T. (2013), *Platformy wielostronne: istota i rozwój relacji z klientami*, "E-mentor", nr 1(48).

eMarketer (2021), *Worldwide Digital Ad Spending 2021*, https://www.emarketer.com/content/worldwide-digital-ad-spending-2021 [accessed: 12.10.2021].

GlobalStats (2021), *Search Engine Market Share Poland*, https://gs.statcounter.com/search-engine-market-share/all/poland/#monthly-202009-202109 [accessed: 12.10.2021].

Grzeszak, J. (2021), *Modele biznesowe mediów po pandemii*, Polski Instytut Ekonomiczny, Warszawa.

IMF (2018), *Measuring the Digital Economy*, https://www.imf.org/en/Publications/Policy-Papers/Issues/2018/04/03/022818-measuring-the-digital-economy [accessed: 29.10.2021].

Interaktywnie.com (2020), *Marketing w wyszukiwarkach*, https://interaktywnie.com/biznes/newsy/biznes/raport-agencje-seo-i-marketing-w-google-dane-analizy-nowosci-porady-ekspertow-260071 [accessed: 12.10.2021].

McFadden, D. (1974), *Conditional Logit Analysis of Qualititative Choice Behaviour*, (in:) Zarembka W.P. (ed.), *Frontiers in Econometrics*, Academic Press, New York.

MIT Technology Review (2016), *The Rise of Data Capital*, http://files.technologyreview.com/whitepapers/MIT_Oracle+Report-The_Rise_of_Data_Capital.pdf?_ga=2.214314772.288411662.1635525199-258656564.1634133834 [accessed: 29.10.2021].

Moazed, A., Johnson, N. (2016), *Modern Monopolies: What It Takes to Dominate the 21st Century Economy*, St. Martin's Press, New York.

Nakamura, L., Samuels, J., Soloveichik, R. (2016), *Valuing 'Free' Media in GDP: An Experimental Approach*, FRB of Philadelphia Working Paper, No. 16-24.

Paliński, M. (2021), *Paying with your data. Privacy tradeoffs in ride-hailing services*, "Applied Economics Letters", https://doi.org/10.1080/13504851.2021.1959891.

Prince, J., Wallsten, S. (2020), *How Much is Privacy Worth Around the World and Across Platforms?*, Technology Policy Institute, Washington.

PwC (2019), *Putting a value on data*, http://pwc.co.uk/issues/data-analytics/insights/putting-value-on-data.html [accessed: 29.10.2021].

SEC (2020a), *Facebook Inc., SEC Fillings Form 10-K Annual Report 2020*, Washington.

SEC (2020b), *Alphabet Inc., SEC Fillings Form 10-K Annual Report 2020*, Washington.

Shah, N. (2015), *You Are Worth $182 To Google, $158 To Facebook And $733 To Amazon!, Arkenea*, https://arkenea.com/blog/big-tech-companies-user-worth/ [accessed: 12.10.2021].

Shapiro, R., Aneja, S. (2019), *Who Owns Americans' Personal Information and What Is It Worth?*, Future Majority, https://assets.futuremajority.org/uploads/report-for-future-majority-on-the-value-of-people-s-personal-data-shapiro-aneja-march-8-2019.pdf [accessed: 12.10.2021].

Shapiro, N. (2019), *What Your Data Is Really Worth to Facebook*, "Washington Monthly", https://washingtonmonthly.com/magazine/july-august-2019/what-your-data-is-really-worth-to-facebook/ [accessed: 12.10.2021].

Sobolewski, M., Paliński, M. (2017), *How much consumers value on-line privacy? Welfare assessment of new data protection regulation (GDPR)*, Working Papers, No. 17, Faculty of Economic Sciences, University of Warsaw, Warsaw.

Srnicek, N. (2017), *Platform capitalism*, Polity, Cambridge.

Śledziewska, K., Włoch, R. (2020), *Gospodarka cyfrowa. Jak nowe technologie zmieniają świat*, WUW, Warszawa.

Thaler, R. (1980), *Toward A Positive Theory Of Consumer Choice*, "Journal of Economic Behavior & Organization", Vol. 1, No. 1.

Train, K., Weeks, M. (2005), *Discrete Choice Models in Preference Space and Willingness-to-Pay Space*, (in:) Scarpa, R., Alberini, A. (ed.), *Applications of Simulation Methods in Environmental and Resource Economics*, The Economics of Non-Market Goods and Resources, Vol 6. Springer, Dordrecht,

https://doi.org/10.1007/1-4020-3684-1_1 .
Winegar, A., Sunstein, C. (2019), *How much Is data privacy worth?: A preliminary investigation*, Harvard Law School, Cambridge, MA.
World Bank (2020a), *Individuals using the Internet*, https://data.worldbank.org/indicator/IT.NET.USER.ZS [accessed: 29.10.2021].
World Bank (2020b), *GDP per capita*, https://data.worldbank.org/indicator/NY.GDP.PCAP.CD [accessed: 29.10.2021].
Yitzhak, Y. (2020), *Here's how much money you made Google by staring at its ads for 20 years*, The Next Web, https://thenextweb.com/news/heres-how-much-money-you-made-google-by-staring-at-its-ads-for-20-years [accessed: 12.10.2021].
Zuboff, S. (2020), *Wiek kapitalizmu inwigilacji*, Zysk i S-ka, Poznań.

Appendix 1. Assessing value using other methods

Our calculations are based on a number of assumptions that affect the final result. To illustrate the scope of the changes in the results, depending on the research method and data sources selected, we analysed the revenue generated by a single user of Facebook and Google using alternative calculation methods. Due to numerous reservations as to the data's quality and adequacy, we only treat it as an auxiliary estimate.

Method	Result	Problem
We took the **values reported in Google Poland and Facebook Poland's financial statements** in the National Court Register as Google and Facebook's income.	Monthly revenue from a single user's data is PLN 0.25 per month per Facebook user and **PLN 1.29 per Google user.**	Underestimating the revenue generated by the data, because, through transactions with companies in Ireland, companies are able to move the place where they report their revenue.
Multiplying the **value of the online advertising market** in Poland by the company's share in this market.	Monthly revenue from a single user's data is **PLN 5.08 for Facebook and PLN 4.43 for Google.**	There is a lack of sufficiently accurate data on Facebook and Google's share in the advertising market in Poland. Additionally, the ad market size estimates do not include foreign advertisers on social media platforms.
Equating the **average time that a user spends on a given platform** with average hourly earnings in Poland.	Monthly revenue from a single Facebook user's data is **PLN 558.92.**	This method inflates the result and is based on the assumption – rarely found in the literature – that time spent on the platform is cognitive work. Moreover, it is difficult to make similar estimates for Google's services.

Source: prepared by PEI.

The estimates of the value that a single user generates for the platform currently available in the literature range from USD 9 (Shapiro, 2019) to USD 13 (Shah, 2015) per month for Facebook, and from USD 3 (Yitzhak, 2020) to USD 15 (Shah, 2015) per month for Google. These amounts are higher than those in our calculations because they refer to American Internet users, who generate more income, or because the calculations take into account market capitalisation, rather than the company's revenues.

Appendix 2. Results of the DCE

We present the full results of the estimation of the parameters of the DCE model below.

MXL Facebook	in WTP-space							
	Means				Standard Deviations			
var.	coef.	sign.	st.err.	p-value	coef.	sign.	st.err.	p-value
Status quo	-1.0018	**	0.3894	0.0101	0.8156		0.5268	0.1216
partial_access	12.3475	***	0.5620	0.0000	32.7547	***	1.3263	0.0000
no_access	17.0706	***	0.7766	0.0000	24.1296	***	1.1226	0.0000
no_profile	3.6017	***	0.4656	0.0000	6.0542	***	0.5451	0.0000
non-personalised_advertising	-1.0416		0.7377	0.1579	8.5599	***	1.0428	0.0000
no_ads	3.8090	***	0.8519	0.0000	5.5890	***	1.0222	0.0000
-Cost	-1.6275	***	0.1248	0.0000	2.9430	***	0.2582	0.0000
Model diagnostics								
LL at convergence	-4620.44							
LL at constant(s) only	-6038.87							
McFadden's pseudo-R^2	0.234884							
Ben-Akiva-Lerman's pseudo-R^2	0.521317							
AIC/n	1.643868							
BIC/n	1.684911							
n (observations)	5664							
r (respondents)	944							
k (parameters)	35							

MXL Google	in WTP-space							
	Means				Standard Deviations			
var.	coef.	sign.	st.err.	p-value	coef.	sign.	st.err.	p-value
Status quo	-2.1657	***	0.4961	0.0000	1.7623	**	0.7246	0.0150
partial_access	6.7145	***	0.6151	0.0000	28.7471	***	1.4970	0.0000
no_access	14.0997	***	0.5579	0.0000	22.5934	***	0.8296	0.0000
no_profile	1.9196	***	0.3848	0.0000	4.6522	***	0.5700	0.0000
non-personalised_advertising	-1.2781	**	0.6417	0.0464	6.2813	***	0.9350	0.0000
no_ads	4.3405	***	0.6289	0.0000	5.1422	***	0.6544	0.0000
-Cost	-1.0969	***	0.1368	0.0000	2.6654	***	0.2350	0.0000
Model diagnostics								
LL at convergence	-4161.89							
LL at constant(s) only	-5898.45							
McFadden's pseudo-R^2	0.294410							
Ben-Akiva-Lerman's pseudo-R^2	0.555585							
AIC/n	1.481953							
BIC/n	1.522996							
n (observations)	5664							
r (respondents)	944							
k (parameters)	35							

Source: prepared by PEI.

List of charts, diagrams, infographics, boxes and tables

The Polish Economic Institute

The Polish Economic Institute is a public economic think tank dating back to 1928. Its research primarily spans macroeconomics, energy and climate, foreign trade, economic foresight, the digital economy and behavioural economics. The Institute provides reports, analyses and recommendations for key areas of the economy and social life in Poland, taking into account the international situation.

Polish
Economic
Institute