"*Your Hashed IP Address: Ubuntu.*" Perspectives on Transparency Tools for Online Advertising

Tobias Urban
urban@internet-sicherheit.de
Institute for Internet Security
Ruhr University Bochum

Martin Degeling
martin.degeling@rub.de
Ruhr University Bochum

Thorsten Holz
thorsten.holz@rub.de
Ruhr University Bochum

Norbert Pohlmann
pohlmann@internet-sicherheit.de
Institute for Internet Security

ABSTRACT

Ad personalization has been criticized in the past for invading privacy, lack of transparency, and improper controls offered to users. Recently, companies started to provide web portals and other means for users to access data collected about them. In this paper, we study these new transparency tools from multiple perspectives using a mixed-methods approach. Still practices of data sharing barely changed until recently when new legislation *required* all companies to grant individual access to personal data stored about them. Using a mixed-methods approach we study the benefits of the new rights for users. First, we analyze transparency tools provided by 22 companies and check whether they follow previous recommendations for usability and user expectations. Based on these insights, we conduct a survey with 490 participants to evaluate three common approaches to disclose data. To complement this user-centric view, we shed light on the design decisions and complexities of transparency in online advertising using an online survey ($n = 24$) and in-person interviews ($n = 8$) with experts from the industry. We find that newly created transparency tools present a variety of information to users, from detailed technical logs to high-level interest segment information. Our results indicate that users do not (yet) know what to learn from the data and mistrust the accuracy of the information shown to them. At the same time, new transparency requirements pose several challenges to an industry that excessively shares data that even they sometimes cannot relate to an individual.

CCS CONCEPTS

• **Security and privacy → Usability in security and privacy**.

KEYWORDS

usability, privacy, transparency , online advertisement, GDPR, SAR

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ACSAC '19, December 9–13, 2019, San Juan, PR, USA
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-7628-0/19/12…$15.00
https://doi.org/10.1145/3359789.3359798

ACM Reference Format:
Tobias Urban, Martin Degeling, Thorsten Holz, and Norbert Pohlmann. 2019. "*Your Hashed IP Address: Ubuntu.*" Perspectives on Transparency Tools for Online Advertising. In *2019 Annual Computer Security Applications Conference (ACSAC '19), December 9–13, 2019, San Juan, PR, USA.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3359789.3359798

1 INTRODUCTION

Advertisements are an essential part of modern online services' business models. A multi-billion dollar industry has evolved around the placement of ad banners and videos that target potential customers [28, 29]. Successful ad campaigns are expected to reach an audience that is likely to be interested in the advertised product and part of a target group defined by multiple attributes like location, age, or interests. To achieve this, ad companies build *behavioral user profiles* which often include data like their assumed interests in products and demographic information as well as clickstream data of websites the users have been tracked on. This personal data is collected or inferred by the ad companies (mostly) without the users' explicit consent or knowledge about these mechanisms [13].

Previous studies have measured users' discomfort with ad personalization [36, 51] and highlighted the importance of transparency as a critical factor [18]. Therefore, scholars have argued that transparency is critical to counter the knowledge imbalance between tracking services and individuals that increments the discomfort [26].

To counter these problems an increasing number of ad-tech companies offer ways to access such data via web portals (e. g., *TripleLift*'s approach [50]) or offer to answer data access requests via email. Through these means, users can gain insights into the data collected about them (e. g., sites they were tracked on) or information inferred from such data. The industry's increase in transparency is likely fostered by new regulation introduced to account for the users' demand to more transparency [25]. The *General Data Protection Regulation* (GDPR) [47] and the upcoming *California Consumer Privacy Act* (CCPA) [46] include the right of each user to request access to the data a company has collected about them (*Right to Access*, Article 15 GDPR).

Prior work on ad transparency only analyzed a small number of services offered by *Facebook* or *Google*, pioneers in this area [5]. The ongoing trend towards more transparency massively extends the number of services that have to disclose information and provide access to user data. In this paper, we present a study on the extent of

new transparency mechanisms and provides insights into how users and companies struggle with new opportunities and regulations.

In a system as complex as online advertising with multiple actors sharing and building upon tracking data, there are multiple challenges to *effective transparency* [35, 40, 57]. First, those collecting the data must be aware of what and whose data they directly or indirectly collect through third parties. Second, transparency is not an end in itself, so when personal data is provided to data subjects, it has to be contextualized and presented in a way that conveys the essential facts but does not overwhelm the user. We study aspects of both challenges by evaluating the current state of transparency tools and the data provided to users when they request access. While we first focus on the views and needs of users, we also try to understand the challenges companies face when providing transparency. In short, we make the following three contributions:

- We analyze 22 transparency tools of online advertising companies regarding their compliance with users' expectations, legal norms, and similar aspects. We find that only three of the tools meet the requirements described in previous work.
- To gain further insights, we conduct an online user study ($n = 490$) to better understand user needs when it comes to transparency in the online advertising ecosystem. We found that—if not explicitly stated—users often do not know who collects their data. Furthermore, users struggle to understand the data provided to them.
- Finally, we investigate the perspective of online advertising companies in an online survey ($n = 24$) and in-person interviews ($n = 8$). They acknowledge problems with existing approaches, some inherent to an ecosystem that is not fully aware of the data flows within.

In summary, our analysis shows that transparency tools do not offer much help to users, yet. It is hard to identify whom someone has to ask for access and often the data is hard to interpret. Participants are not able to draw conclusions. This is reflected in the low number of requests that are reported by industry experts. Going forward tools need to be developed that help users better understand what data is used for what purposes, while at the same time the industry has to find adequate means to communicate what data they are collecting (sometimes it seems that have to learn that for themselves first).

2 BACKGROUND

Many digital services such as websites or mobile apps rely on revenue from displaying ads to their customers. In 2017, the online ad industry generated an estimated revenue of over € 41.8 billion [29] in Europe and $88.0 billion US dollars in the US [28].

Online Behavioral Advertising (OBA). OBA is a technique to tailor ads to individuals based on their online behavior, on their clickstream [9], or other personal data like IP addresses. To perform ad personalization, companies need to collect data, often by tracking users across the web or utilizing a service that does that, and therefore user tracking has become an essential part of the business model of web services [45]. Unique identifiers are assigned to each user, either generated by the ad company or computed based on properties of the users' device (so-called *device fingerprinting*) [20]. In the mobile world, unique *advertising identifiers* are used to identify users. These identifiers are often provided by the operating system of the phone and are only accessible from apps installed on the phone but not from web pages [56]. Similar to cookie deletion and opt-out mechanisms in web browsers, users can choose to reset these IDs or turn them off altogether, preventing companies from recognizing them [32].

Available ad space is sold on *real-time bidding* (RTB) platforms whenever a user visits a website. Different entities are involved in the RTB process, but the general flow of information, as described by Yuan et al. [58], is as follows: When a user visits a website, the site provides the available ad space (formally called inventory or impressions) to an ad exchange service which starts auctions for the available impressions on the site. Websites often use a service (so-called *supply-side platforms*) to provide the inventory. Now, several *demand-side platforms*, who help to manage ad campaigns, place bids on the ad space depending on their estimated value of the impression. These bids are placed on behalf of the advertisers (e. g., brands) who want to place ads. The highest bid wins the impression, ensuring that the ad selling price is maximized.

Legal Background. In 2016, the European Union (EU) harmonized data protection laws of its member states by introducing the *General Data Protection Regulation* (GDPR or Regulation 2016/679) [47], which went into effect on May 25, 2018. Among other legal obligations, the GDPR specifies under which circumstances the personal data of EU citizens may be processed and defines the obligations of companies processing the data. Similarly, the *California Consumer Privacy Act* (CCPA) of 2018 [46] aims to strengthen privacy rights of California residents. The CCPA is planned to go into effect on January 01, 2020.

Article 15 GDPR describes an individual's *right to access*. The right to access describes which information data controllers have to provide to users upon request. This includes information typically found in a privacy policy like the categories of personal data processed, the purpose of processing, or the right to file a complaint with data protection authorities. In addition, Article 15 grants users the right to access their personal data (*"The data subject shall have the right to obtain [...] access to the personal data [...]"*). Article 20 GDPR extends the access right to the *right to data portability*, meaning that an individual may not only review data stored about them but can also request a copy of the collected personal data. Similar to the GDPR, the CCPA requires that starting in 2020 *"a business that receives a verifiable consumer request from a consumer to access personal information shall promptly take steps to disclose and deliver, free of charge to the consumer [...]"*. Requests to data access are referred to as *subject access requests* (SAR).

3 ANALYSIS OF TRANSPARENCY TOOLS

Some ad-tech companies implemented ways to give individuals access to data stored about them to account for growing user demand. Most notable, *Google* and *Facebook* developed privacy dashboards or transparency portals after their data collection practices had come under public scrutiny [5]. Other businesses have also set up information sites or web forms individuals can use to request their data. Recently, the number of available tools has grown to account for regulatory obligations of the GDPR and CCPA that require

companies to give individuals access to collected personal data. Several examples of data such tools provide to users are shown in Figures 3, 4, and 5 in Appendix A.

Companies to Analyze. Previous work has shown how difficult it can be to get access to data and that requests not always successful [52]. To avoid overhead of tedious and possibly unsuccessful access requests, we looked at all members of large online advertising alliances (i. e., the *Network Advertising Initiative* (NAI), the *Interactive Advertising Bureau* (IAB), and the *Digital Advertising Alliance* (DAA)) and checked which company offered an online tool to access personal data. If a company offered an online tool, we analyzed it in our study.

According to public statements of these alliances, they represent over 5,500 companies. However, they have only 500–600 (distinct) members listed on the organizations' websites which we all manually reviewed. We analyzed all online tools we found and asked for access to our personal data with companies that do not provide an online tool but did reportedly grant access in an easy way. In total, we analyzed 22 web portals (15) and responses (7) to our subject access requests. These companies include ad-tech industry giants (e. g., *Google* and *Facebook*) and medium size companies (e. g., *Sojern* and *MediaMath*). We differentiate between two types of how users can access their personal data: *online* and *offline*. By *online* we mean that users can visit a website which (automatically) reads the user's cookie store and shows personal data associated with the read identifier. If the data is provided in a file format (e. g., `.csv` or `.pdf` files), we labeled it *offline*. Using a VPN service (US–VA), we also verified that all online tools are also available to US-based IP addresses.

3.1 Criteria Definition

We evaluated the transparency tools based on heuristics from multiple sources: (1) user expectations elicited in previous studies, (2) descriptive information found in the privacy policies, and (3) self-regulative norms proposed by industry groups. In the following, we describe all criteria we used to analyze the transparency tools of the 22 companies in detail.

User Expectations. Previous work has shown that users have different—mostly negative—views on online behavioral advertising (OBA) but also demonstrate a need for transparency in OBA. In the following, we describe criteria found by previous work to be most important to users when it comes to transparency and understanding of OBA.

Interest segments/Demographics: Dolin et al. found that users are more comfortable with OBA if they are aware of the connection between the created profile and their interests [18] (criterion C1). They also found that users' comfort with personalized ads is (positively) correlated with the perceived sensitivity of the data category (e. g., health-related information is seen as critical). One approach to increase users' comfort could be to show the segment data assigned to the user by a company, although previous work has shown that these profiles tend to be inaccurate [5, 42]. Besides assumed interests, OBA is often based on (inferred) demographic information (e. g., ethnicity, age, location, or salary). Discrimination based on this demographic information is a big concern [41] (criterion C2). Displaying interest segments or demographic data to users can help them understand how companies use the collected data and why they see specific ads.

Tracking and Clickstream Data: Ur et al. found that users' views on online tracking can range from "useful" to "scary" [51], and other work highlights the dislike of ads based on clickstream data [37]. Therefore, when companies disclose on what websites they have tracked users, it can be helpful to understand *why* certain ads are displayed, e. g., in cases of re-targeting where ads are based on products a user has previously viewed online. As shown in Figure 5 (App. A), clickstream data is made accessible in varying granularity ranging from raw data that includes user-agent and other technical information to lists of timestamps and websites recorded.

The willingness of users to share data with advertisers was analyzed by Chiasson et al. [11] (criterion C3). They found different factors that influence the willingness to share and show that users have complex privacy needs which are not accounted for by current tools. Our work differs from the named approaches as we measure the *transparency needs* of users concerning OBA instead of analyzing their attitude towards different aspects of OBA.

To summarize, previous work found three criteria users expect to find when evaluating personal data an ad company collected: (1) interest segment data [18], (2) demographic data [41], and (3) tracking data [11]. We inspected the data provided by the tools of the 22 companies and checked if the data can be grouped into any of these groups and did also check if the privacy policy did state if such data was collected/inferred.

Privacy Policies. The content of privacy policies should be helpful to users who want to learn more about the privacy practices of a company, aside from being compliant with legal requirements. Therefore, we analyze the privacy policies of the 22 companies regarding three criteria: (C1) Does the policy use plain and clear language, (C2) Is the purpose of data collection explained, and (C3) Is the way how data is collected explained?. Thus, criteria C2 and C3 focus on the requirements for technical descriptions, not compliance in general.

To assess the readability of the privacy policies (C1), we used the Flesch-Kincaid Grade Level (FKG). This grading assesses how many years of school education one needs to understand a text (e. g., FKG = 12 means 12 years of education—senior-level high school student in the US). There is no consistent usage of readability scores in the literature, but Fabian et al. [23] found that the FKG, among other scores such as SMOG, RIX, or LIX, produces comparable results between each other (i. e., the correlation coefficients are almost 1). Previous work often does not report how they actually calculate the score (e. g., [23, 30, 31]) and we found that different available tools compute different FKG scores for the same text because they compute sentence endings or syllables differently. We computed the FKG score using the *koRpus* R package [39].

Regarding criteria C2 and C3, the privacy policies of the companies were independently analyzed by two researchers with experience in the area to check whether the required explanations were given. All researchers have a strong background in data protection, a strong understanding of the online ad ecosystem, and experience in the field. One is also a certified data protection officer with legal

expertise. We told these researchers that a technical description, although it might not be understandable by most users, is sufficient. While this favors the companies, we assume that if users are interested in how or why data is collected, they could check what these technologies are and how they work. While it would be favorable, it is—from our point of view—not the purpose of a privacy policy to explain technical details of the used technologies.

Industry Self-Regulation. The OBA industry associations have developed transparency guidelines for their members on how and which kind of information and choices they should provide to consumers (e. g., the DAA and IAB [16, 21]). As noted earlier, we analyzed the guidelines of the three most prominent alliances, the *Interactive Advertising Bureau* (IAB), the *Network Advertising Initiative* (NAI), and the *Digital Advertising Alliance* (DAA). All guidelines urge companies to take steps to increase transparency to their users and every company we analyzed is a member of at least one of the alliances. However, we found that the guidelines are quite vague and not easy to validate on the users' end. For example, companies are asked to place a special icon (which provides a link to an opt-out tool and further explanations why the ad is displayed) on the ad if it is based on a behavioral profile. If the ad banner does not contain such an icon, this could mean that the company either provides an ad not based on a profile (e. g., depending on the website's content) or that it does not adhere to the self-imposed rules. By manual inspection of several websites, we found the same ad twice once labeled with the icon and the other without the label. It is possible that this observation was coincidental and that one ad was contextual and the other was profile-based, but this illustrates that it is nearly impossible to decide if the guidelines were followed or not. Due to this inconsistency and previous work highlighting the ineffectiveness of such icons [43], we did not further investigate this transparency mechanism.

Besides the guidelines, the DAA, NAI, and IAB provide and maintain websites for consumers to learn more about online advertising and control privacy related settings (e. g., mechanisms to opt-out of OBA for several ad companies at once). The DAA provides the "YourAdChoices" [17] and the EDAA the "Your Online Choices" [22] tool. At the time of this study, none of the guidelines contained rules or advice how users can obtain access to their personal data.

3.2 Results

Table 1 lists the results of our analysis of transparency tools in alphabetic order. Ten companies provided data online only, eight companies provided data offline only, and five companies provided data in both ways. In general, online data can be seen as more usable since it is pre-processed, while most ($n = 6$) of the offline data is comma-separated, which needs a more technical background to interpret. The table lists how users can access their data (*Access*), which data is provided (*Expect.*), which data is provided in the company's privacy policy (*Privacy P.*), and further information that help users to understand the company's usage of personal data (*Misc.*). The category *Misc* lists if the company provides insights with whom personal data was shared (*Sharing*) and whether the used technologies to track users are explained. Some profiles contained inconclusive information (e. g., Segment: `companyB_Usersync_Global` or `Your hashed IP address: Ubuntu`). If we could see (or guess) the meaning of such information, we ruled in favor of the companies.

Table 1: Results of transparency tools analysis. *Access* describes the format how users get access to their data. *Expect* shows whether the provided data contains information to our defined categories. *Privacy P.* lists whether privacy policies are useful to users. *Misc.* lists additionally provided data of interest. ○: Does not apply. ●: Applies according to the privacy policy and data is provided. ◑: Applies according to the privacy policy but no data is provided. †: *Google* and *Facebook* only shows tracking data on their own platforms. *Twitter*'s way to provide sharing data did not work for us. *Sovrn* only shared pseudonyms of partners. ϟ: Our analyzed profile did not include this data but could include it. ◇: These companies only shared their syncing partners.

Company	Access: Online	Access: Offline	Expect.: Segments	Expect.: Demographics	Expect.: Tracking	Privacy P.: FKG	Privacy P.: explain why	Privacy P.: explain what	Misc.: Sharing	Misc.: explain tech.
Adform	✗	✓	●	◑	●	13.40	✓	✓	✗	✓
Amobee	✓	✗	●	◑ϟ	◑	13.05	✓	✗	✗	✗
AppNexus	✓	✗	●	●	◑	11.96	✓	✓	✓	✓
ComScore	✓	✗	◑	●	●	10.63	✓	✓	✗	✗
Conversant	✓	✓	●	◑ϟ	●	13.43	✗	✗	✓	✗
Criteo	✗	✓	◑	◑	●	12.00	✓	✓	✓	✗
Facebook	✓	✓	○	●	◑†	14.19	✓	✓	✗	✓
Google	✓	✓	●	●	◑†	13.37	✓	✓	✗	✓
Leiki	✓	✗	●	○	●	9.96	✗	✗	✗	✗
Lotame	✓	✗	●	●	◑	12.72	✓	✓	✗	✓
MediaMath	✗	✓	●	●	◑	12.69	✓	✓	✗	✓
Oracle	✓	✗	●	●	◑	11.92	✓	✗	✗	✗
Quantcast	✗	✓	●	●	●	12.10	✓	✓	✗	✓
Rubicon Project	✓	✗	○	○	●	12.77	✓	✓	✓	✓
ShareThrough◇	✗	✓	◑	◑	◑	12.07	✓	✓	✓	✗
Sizmek	✓	✓	◑	●	◑	14.81	✗	✓	✗	✗
Sojern◇	✗	✓	◑	◑	◑	13.96	✓	✓	✗	✓
Sovrn◇	✗	✓	◑	◑	◑	14.24	✓	✗	✓†	✓
SpotXchange	✓	✗	○	●	◑	12.15	✗	✗	✓	✓
The Trade Desk	✓	✗	●	◑ϟ	◑	10.29	✓	✓	✓	✗
TripleLift	✓	✗	●	◑ϟ	◑	12.19	✓	✓	✓	✗
Twitter	✓	✓	●	●	○	12.16	✗	✓	✗†	✓

User Expectations. We checked if the transparency tools provide data on three levels, as described above. These categories contain information the ad companies inferred/collected about the user regarding (1) interest segments, (2) demographics, and (3) tracking. We inspected the data provided by the tools, checked if the data can be grouped into any of these groups, and also reviewed if the privacy policy stated if such data was collected/inferred.

We identified three different cases: (1) a company states that they collect data in one of the categories and provides this data (●), (2) a

company states that they collect data on one of these categories but does not provide such information (◑), and (3) the company does not state that they collect data in one category and also does not provide any information (○). We did not observe the case that a company did not state that they would collect data on a category but provided such information. In some cases, the 22 analyzed profiles did not contain interest segment or demographic data. However, the profiles might contain such data for other profiles because the shared data is not a full set of all categories, but only the data they assigned to one user. Meaning, that a company might try to infer for example the user's age but could not do so based on the collected data. We found four cases where this applies (marked with ϟ in Table 1).

Thirteen companies provided information regarding inferred segments and six companies chose not to provide this data, despite the fact that the privacy policy mentions that segments are inferred. Nine companies provided demographic information they inferred from the users' online profiles we analyzed. If the companies provided access to demographic information, most of them shared the user's location(s) or the inferred age. In general, users do not think that companies trying to learn their age is a severe privacy problem [41]. Two companies provided health-related information (e. g., *Health & Fitness > Diets & Nutrition*). No company provided racial information, which is in line with most privacy policies stating such information is not inferred.

While most companies state in their privacy policy that they track users around the web, we only found six companies that list the websites on which they tracked the user. *Google* only provided the visits to websites the company controls (e. g., *YouTube*), while it is known that they track users across tons of sites [20].

Legal Requirements. Two researchers—both with a strong understanding of the online ad ecosystem—were assigned the task to classify if companies disclose "why" and "what" data is collected. The final inter-rater agreement for classification shows substantial agreement (Cohen's $\kappa = 0.77$ for "why" and $\kappa = 0.74$ for "what" data is collected; agreement > 90 % for both categories). In the rare cases of discrepancies, a third person with a similar background was consulted to resolve such cases. We found that five companies do not disclose *why* and five companies do not disclose *what* they collect data. It is worth noting that some companies only vaguely explain why they collect data (e. g., *<Company Name> "advertising technology allows Business Partners to target advertisements to users [...]"*) or how this is actually implemented. In three cases neither *why* nor *what* data is collected is given.

We computed the Flesch-Kincaid grade for all privacy policies. The score suggests that on average users need 12.58 (with SD 1.21) years of education (senior high school student in the US). Data provided by the *U.S. Census Bureau* shows that 12 % of all adult US citizens did not obtain a high school diploma [54].

Summary. The field of analyzed transparency tools is heterogeneous with regard to the type of data provided, the way of access, and information about sharing activities. Some companies' reports lack of explanations of data collection and usage, and they do not provide all data representations that they claim to have. As the tools provided data in different forms and levels of granularity, it is worth analyzing to which extent such data helps users to assess the privacy implications of a company's activities.

4 PERCEPTION OF TRANSPARENCY TOOLS

Previous work has focused on the transparency of targeted ads themselves [43] or the accuracy of inferred interest segments [5, 42]. In this study, we try to get a better understanding of the users' expectations and needs when it comes to transparency in online advertisement.

4.1 Method

To get a better understanding of the users' side of transparency tools, we run an online survey which focuses on two aspects, First, we wanted to understand to what extent users can identify who is collecting their data because otherwise, they would not be able to request it. Second, the ways companies provide access to data differs from approaches studied in the past. Our goal was to understand how different types of data disclosures found in the field help participants understand the privacy implications of a company.

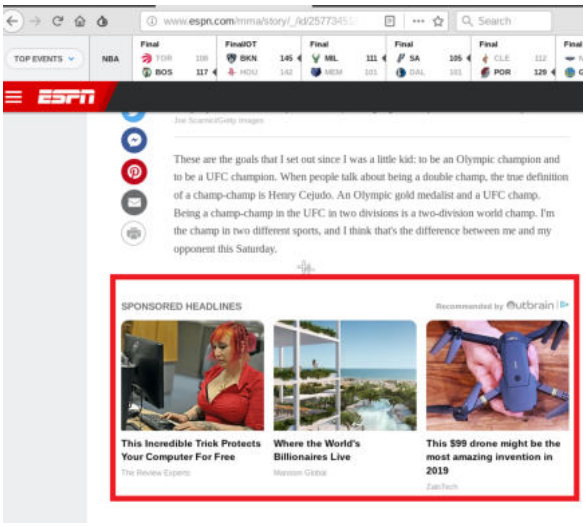
Our study focuses on the ability of users to understand the provided data as it is the most important mechanism to provide transparency, while other aspects like completeness or the comprehensibility of how certain information was inferred also play an important role in the value of these tools.

Prior to the analysis of users' perceptions (see Appendix A), we asked questions about their attitude and understanding of online advertisements. We also asked them about their general view and usage of the tools to get access to personal data.

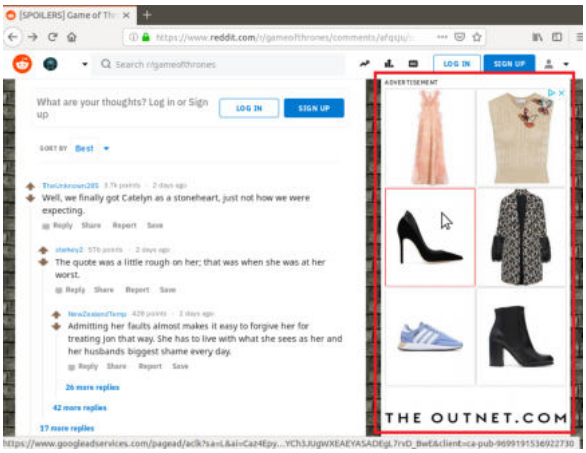
To test if users can identify which ad network is responsible for an advertisement, we present two screenshots of websites, one of which contained a standard ad banner(see Fig. 1a—top image in Fig 1) and the other an advertisement with links to articles distinguished as "Recommendations" (see Fig. 1b—bottom image in Fig 1). The ad banner contains a link to an opt-out program but does not directly show the name of the third party. Users would have to hover over the ad with the mouse and check the URL displayed in the browser's status bar to identify the ad network—we included this URL in the screenshot but did *not* highlight it. The recommendation contains a reference to the third party that generated the ad (i. e., *"Recommended by Outbrain"*).

The remainder of the survey focuses on the users' expectations and understanding of personal data provided by ad companies upon request. To assess this, we took screenshots of nine real-world profiles that were provided to us upon request.

We grouped the nine profiles into three categories based on their content: (1) technical data, (2) tracking data, and (3) segment data—a definition of these categories can be found in Section 4.2. The order of these categories was randomized for each participant. To reduce the length of the survey, we did not differentiate between segment and demographic data. The influence of this data on the perception of online ads has been studied before [41]. Instead, we differentiated between more abstract clickstream (tracking) and detailed technical data. Our analysis of existing transparency tools (see Section 3) showed that disclosing data in this form is common. The disclosure of raw data is also related to the new right to data



(a) Article recommendation including the company providing it (top right corner)



(b) Traditional ad banner

Figure 1: Article recommendation (top) and ad banner (bottom (red frames) referenced in Q10 and Q11 (see Appendix A). The article recommendation discloses the company providing it.

portability and the perception of this form of data for transparency has not yet been studied.

Each section of categories starts with a brief introduction to the data shown, followed by three different examples of profiles. Participants had to assess four statements regarding their understanding and the presentation of the data. For most questions, we used 5-point Likert scales and, for the remainder, we used "Yes/No" questions and a prioritization question (see Appendix A). We provide a 5-point Likert scale (ranging from *"Strongly Agree"* to *"Strongly Disagree"*) and an *"I prefer not to answer"* answer option for each statement. The order of the profiles is also randomized within each category. At the end of each section of categories, we asked the participants if these profiles would help them to better assess the privacy impact of the companies. In our survey, all questions (aside from optional open-ended questions) provide a *"I prefer not to answer."* answer option. Following each section of categories, we asked participants general questions regarding their personal views on provided data and preferences which data representation and the category they prefer.

In general, we used Pearson's chi-squared test to test the independence of two variables and the Pearson correlation coefficient to determine a linear correlation between two variables. For both tests, we used a significance level of $\alpha = 0.5$. Furthermore, we assigned the value 5 to the most positive answer in a Likert scale, 1 to the most negative answer option, and consequentially three to the natural option (e. g., "Strongly Agree" = 5, "Undecided" = 3, and "Strongly Disagree" = 1).

We conducted a pre-study ($n = 50$) with a similar survey structure as described above. In the pre-study, we focused on how users might use data access to their benefit (i. e., how they would use the provided information). However, due to usability problems, users did not give useful feedback on this (e. g., P-6 stated: *"No, this is gibberish to me"*). Thus, we dropped this question for the final study. The full questionnaire is shown in Appendix A.

To recruit participants, we used Amazon's *Mechanical Turk* (MTurk) [3] and only accepted participants with high task completion rates ($\geq$ 97 %) and permanent residents of the US. Furthermore, we only accepted participants who were at least 18 years old and asked for their consent to participate in the survey. In the introduction of the survey, we disclosed our names, affiliations, and all sponsors. We used a self-hosted LimeSurvey [34] instance to conduct the survey. Participants received 2 USD for completing the survey which took them around 15 minutes on average (median = 13 min). All answers were saved pseudonymously using the random unique string, used by MTurk to pay the workers. After the payment process with MTurk had been completed, we deleted the identifier to increase participants' anonymity level.

4.2 Results

The survey was conducted in February 2019 with 490 participants, using Amazon's Mechanical Turk to recruit them. and in the following we describe the main results.

Participant Demographics. 54 % of the participants are male, the majority (46 %) of participants are between 25 and 34 years old, and holds either a high school diploma (33 %) or a bachelor's degree (52 %). The full demographic information in our study, compared to the general adult US population provided by the *U.S. Census Bureau* [54], can be found in Table 2. Our sample is biased as more participants identified as male, have a better formal education, and are younger than the general population. A majority of participants (90 %) use some form of privacy protection online, at least from time to time (ad blocker: 50 %; private browsing: 52 %; delete cookies: 71 %; opt-out: 31 %; none: 10 %). A recent study found that 37 % of Internet users use an ad blocker, especially younger individuals [19]. We assume that we observed more ad blocker usage since our sample is skewed towards younger participants.

Table 2: Participant Demographics. ★: The census data does not account for non-binary individuals. ♣: The census data combines these categories.

	Amount	%	US pop.
Gender			
Male	264	54 %	49 %
Female	224	46 %	51 %
Non-binary	1	0 %	— ★
Age			
18–24	41	8 %	16 %
25–34	218	46 %	22 %
35–44	112	23 %	20 %
45–54	66	14 %	21 %
55–64	40	8 %	21 %
Education			
None	1	0 %	12 %
High School	161	33 %	51 %
Bachelor's	255	52 %	18 %
Pro./Master's/Ph.D.	69	14 %	11 %♣

Attitude towards online advertising. The general view of participants on online advertising is quite neutral. Participants who see ads that suit their interests still evaluate them slightly negatively (mean: 2.7 with SD 0.4 and the hypothesis test yielded $p \leq .0005$), which is in line with previous work [18] (Q1 and Q2). Other studies also found that users find ads "creepy" or "intrusive" [51]. In our study users expressed such views too, but at the same time they did not evaluate ads negatively (e. g., P-02 stated in Q9: *"They [ads] are creepy, a product is merely mentioned in my house then I see ads for it the next time I'm online"* but at the same time stated her views on ads as "Moderately satisfied").

The neutral view on online ads is likewise observed in an open-ended question (Q9) and on a Likert scale (85 % of participants choose one the following answer options almost balanced (Q2): "Moderately satisfied" (= 4), "Neither satisfied nor dissatisfied" (= 3), or "Moderately dissatisfied" (= 2) with a mean of 3.07 and SD 0.5 (Q. In total, 73 % of participants *"agreed"* (47 %) or *"strongly agreed"* (26 %) with the statement that access to personal data is useful to assess privacy implications of the usage of their data (Q5), but only 19 participants requested their data, mostly from big companies like *Google* or *Facebook* (Q4).

In Q6, 60 % of participants stated they were *"not"* (10 %) or *"somewhat knowledgeable"* (50 %) about online advertisements while 30 % stated that they were *"very knowledgeable"* (5 %) or *"knowledgeable"* (25 %). We used a multiple-choice question to test this self-assessment (Q7). This question contained seven statements on the online advertising industry—four of which are correct and three incorrect. Each (multiple-choice) answer option was selected 357 times on average (SD 51) regardless of whether it was true or false. This indicates that users often do not understand online advertising as well as they think they do. Furthermore, some participants made statements containing false information about online advertising: *"I honestly expect some ad companies to illegally collect my facial expressions and sounds in my environment through cameras and microphones. I always expect them to access other apps and histories of everything that I do"* (P-152). While there was no public report on this happening in practice at the time of the survey, after finishing the study, multiple reports surfaced that big Internet companies employed humans to categories conversations recorded by smart home devices [12, 27, 49]. The most common misconceptions were that ad companies have access to *all* purchased products (64 %) and the *full* browsing history (66 %).

65 % of all participants stated (Q22) they thought that companies did *not* provide all collected information upon request, only 13 % thought they would do that, and 22 % had no opinion on this topic. This mistrust might be based on misconceptions about what companies can collect or relate to public reporting on data leakage scandals (e. g., *Facebook* providing data to *Cambridge Analytica* and *not* informing users properly [48]).

Identifying Data Collectors. We tried to assess if users can understand what personal data is used when they see a specific ad and whom they have to ask to get access to this data (Q10 and Q11). To do so, we showed users two screenshots of websites containing ads (see Figure 1 in Appendix A); one contained information regarding the company providing the ad, while the other did not. We asked users whom they would have to contact if they wanted to understand *why* they see this specific ad. We provided different answer options (multiple choice): (1) the website on which the ad was shown, (2) the company name of the advertised product, and (3) the actual ad network providing the ad.

In the case of the ad that contained (Q11) the ad network's name, 46 % of users answered the question correctly, 28 % named the advertised company, 17 % named the visited website, and the remaining 9 % did not know whom they have to ask. For the ad banner that did *not* directly include the ad network's name but showed it in the link when hovering over the ad (Q10), 43 % named the advertised company as the contact company, 24 % named the visited website, and only 24 % correctly knew whom they should have to contact. In conclusion, only a minority of participants was able to identify the correct company to contact, but we did not find a significant correlation between users' self-assessed knowledge about online advertising and their answers (Pearson-correlation: $r = 0.56$ with $p = 0.57$). This shows that users have trouble identifying who is collecting their personal data when it comes to ads, but stating the name of the collecting company close to the banner helps.

Assessment of provided data. In the survey, participants are shown three data category sections, each listing different data types we received by using the different transparency tools (Q12–Q21). The categories are (1) technical data, (2) tracking data, and (3) segment data. *Technical data* is raw data presented to users without any kind of processing, typically in a text file. This is likely log data that can be directly extracted from HTTP traffic (e. g., user agents) or directly derived from this data (e. g., locations based on IP addresses). *Tracking data*, also often provided in text files, is information on websites on which the ad company has tracked the user (clickstream data)

or with whom personal data was shared. (Interest) *segment data* is information ad companies inferred from a user's online behavior.

The results of participants' views on the provided data are shown in Figure 2a. Tools that shared inferred data (i. e., segments) were evaluated very positively in all four question categories. Over three quarters (76 %) of participants stated (sum of answer options "*Strongly Agree*" and "*Agree*") that they understood the provided data and thought that it was helpful to understand what companies do with personal data. It is worth noting that not all companies infer this kind of data. For example, a company offering a service to identify ad fraud will most likely not infer high-level data but still has access to personal data like IP addresses. Tracking information is less easily understood by participants: Close to half of the participants (53 %) report that they understood the data and found it helpful (47 %). Profiles that provide technical data were rated slightly less understandable but much less helpful (39 %), and in these profiles, data is presented in a much less clear way (37 %) than in the other categories of profile. We found a correlation between all four questions on clear presentation, understandability, helpfulness, and whether users anticipated this type of data, in each section of categories (Pearson correlation: $r > 0.5$, $p < 0.001$). Meaning that all four questions in each category were answered similarly.

After presenting all three profiles of one category, we asked participants if such data is useful to assess the privacy impact of companies. The results are given in Figure 2b. Similar to the assessment of profile categories, segment data is rated to be most helpful, followed by tracking data. When it comes to preferences which data users would like to receive when they performed an access request (Q23), participants equally rank "tracking data" and "interest data" as first choice (41 % for tracking and 45 % for segment data) but more users chose "tracking data" as their second choice (47 % vs. 28 %—see Figure 2c). This is unexpected as participants stated they found segment data to be most useful (see Figure 2b) to assess privacy implications of a company, and one would expect that they also prioritized profiles accordingly. While participants did not directly comment on this discrepancy one explanation could be that tracking information is accurate and factual, while segment information is inferred and previous research has shown that it is often inaccurate [42]. Therefore, segment data helps users to understand why they see specific ads but tracking data gives them a better idea of what data companies have. In general, combined overall profile types, participants who stated that they understood the provided personal data thought that it was useful to assess the use of data ($p < 0.0001$) and stated it was presented transparently ($p < 0.0001$). Furthermore, participants who stated that the presented data was useful—regarding the usage of data—also stated that the data helped to bring more transparency to the advertisement ecosystem ($p \approx 0.0005$). Access to technical data is by far the least favorite way how users want to get access to their personal data. However, technically savvy users who answered our question regarding the ecosystem (Q7) correctly and stated that they were at least "*knowledgeable*" (Q6) about the online tracking industry rated this kind of information more helpful ($p < 0.001$).

When asked (Q24), 55 % of participants expressed that, after seeing personal data collected on a stranger, that they were *"very interested"* or *"interested"* in data collected about themselves. 58 % stated that they would change their online behavior due to the seen data (Q25). Considering that only a few had requested their data previously, this could be related to a social desirability bias, but it could also indicate that there is simply a lack of awareness of transparency tools.

Summary. The analysis of users' perception of available transparency tools shows that individuals prefer the collected data being shown as interest segment information over raw technical data. Furthermore, while participants were aware of the privacy impact of ad companies, only very few did request access to their data but reported they might do so.

5 BUSINESS PERSPECTIVE

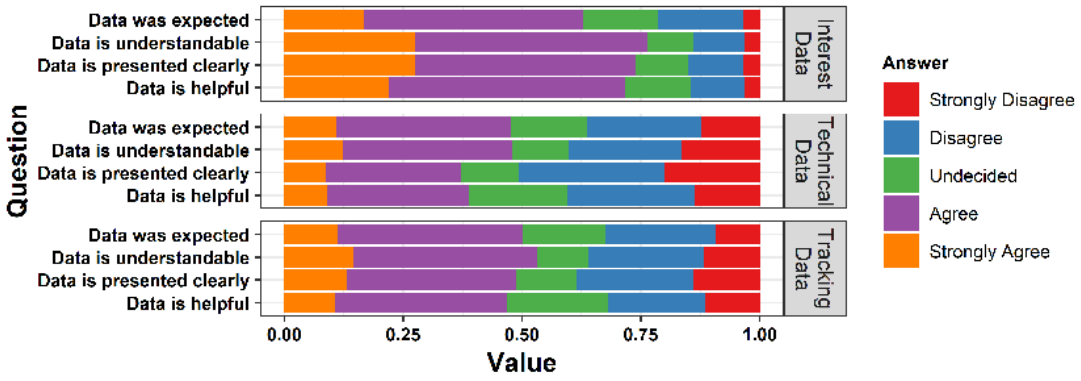
Aside from evaluating existing transparency tools, we contacted 773 companies to assess their view on the usefulness of transparency tools for users. We motivated our inquires not purely on transparency tools in general but also wanted to assess how current legislation (i. e., the GDPR) influenced the design of such tools and which challenges companies faced. To this end, we crawled 773 privacy policies of the companies we analyzed and extracted all email addresses present in the policies, using a regular expression. By manual inspection, we identified the relevant email addresses by name (e. g., `privacy@company.com`) and dropped other addresses (e. g., `sales@company.com`) if we found a more specific one. We did not find an email address in 35 policies via automatic or manual inspection. We used the regular email server from a research organization to send the emails.

5.1 Method

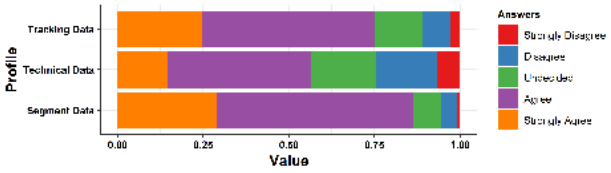
We sent the first batch of email invitations to 74 companies (approx. 10 %) in October 2018. In December 2018, we invited 333 companies (approx. 45 %) to participate in our survey. In January 2019, the final batch of 333 companies was invited. Out of all invitations, 147 resulted in an error message. 48 responses said that our email had been classified as spam/batch mail, the rest resulted in other delivery errors. We manually double checked all of these email addresses for transcription errors. While one company did not exist anymore, the other companies listed email addresses in their privacy policies that do not exist at all. Overall, 593 companies received an invitation to participate in our study.

The survey consists of four categories of questions: (1) general questions regarding the company's business processes for "*subject access requests*" (SAR), (2) four questions regarding their development of SAR processes, (3) three open-ended questions regarding the company's view on the usefulness of data transparency and the GDPR in general, and (4) demographic questions. The full questionnaire can be found in Appendix B. At the end of the questionnaire, participants were asked to volunteer for an in-person interview.

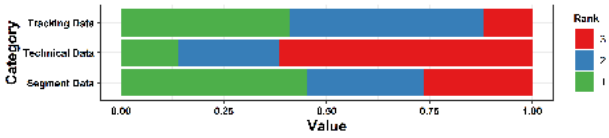
Interviews were conducted using a semi-structured guideline as described by Flick [24]. After a brief introduction, interviewees were asked to explain their job position, if applicable, the business model of the company they are working for, and then give a broad assessment of what the GDPR meant for their business. Afterwards, the interviews focused on the same topics as the survey, but we also asked participants for their personal opinion about how to improve transparency. The interviews were conducted in January and February 2019. All companies that completed the survey and



(a) Evaluation of different aspects of the provided data.



(b) Participants' view on the usefulness of profile categories.



(c) Participants' ranking of profiles.

Figure 2: Assessments of different variants to display personal data.

participated in the interviews did so on the request of anonymity and are not necessarily the companies whose privacy tools we analyzed in Section 3. Interviews were transcribed focusing on content and analyzed using the thematic analysis method [8]. After initial coding to structure the data it was structured in themes discussed by multiple interviewees' to identify their positions. In two cases the analysis was based on notes because one participant did not agree to be recorded and in one case the recording was of bad quality. The coding and analysis was performed by one of the authors.

5.2 Results

Of the 593 invited companies, 24 (4 %) completed the survey and eight agreed to participate in an in-person telephone interview. In 14 cases, the survey was taken by a company's Data Protection Officer or a person from the legal department. In the remaining cases, we got responses from persons on the executive level in the company (e. g., "head of data business" or CEO). Twelve of the participating companies have their headquarter within the EU, nine outside the EU (eight in the US and one in Israel), and two companies preferred not to disclose this information. In nine surveys it was reported that the company employs 1–100 people, ten had 101–500, and four companies had more than 1,000 employees. The size of the department responsible for handing privacy-related requests is in all but two cases proportional to the company size (i. e., larger companies have larger privacy departments). The companies where this is disproportional are companies specialized in privacy-related consulting.

Of the eight interview partners, four companies directly collect and process personal data, two companies only handle personal data on a business-to-business side, and two companies handle both. All stated that they handle personal data of at least one million data subjects. Two of the interviewees were privacy consultants and claimed to work for *"dozens"* of clients in the ad industry. In the interviews, they answered the questions generalized for all of their clients. We report all findings and quotes in the singular to provide the same level of anonymity to all interview partners.

Impact of the GDPR. In the interviews, several partners highlighted that—from their point of view— achieving GDPR compliance in online advertisement is one of hardest tasks (I-2: "*So I represented advertising in the GDPR groups, so we kind of created a group of champions, if you want, so I was the champion on advertising, and it proved that advertising was the most challenging one [...]*"). Not surprisingly, all participants stated (in the interviews as well as the survey) that the GDPR helped them to convince their management to invest in privacy. However, aside from these increased costs the only other benefit mentioned by four companies was that GDPR compliance might be a competitive factor (e. g., P-17 stated: "*It became a marketing talking point that we used to show that people want some control over what data is collected about them, but other than that it has had almost zero benefit and quite a bit of cost*").

When asked about the impact of the GDPR on their daily work we got mixed responses. Some companies stated—aside from the two months prior to the GDPR—that the workflows in their companies did not changed much while others reported problems that, as a result of the GDPR, data they had considered non-PII now is considered PII data (e. g., I-4: "*[...] suddenly, any identificator* '[sic.]

would become PII. [...] *they* (the GDPR) *say that cookies are personal identifiable information.* [...] *So suddenly everyone, absolutely everyone working in advertising, would handle PII data*"). "PII" is the term used in the legal debate in the US. Interview partners that claimed their processes had already been GDPR-compliant beforehand stated that they had to add modules that implement the required accountability (e. g., records when and how consent was conceived). One company stated that they had eliminated all personal identifiers not necessary for their core business, a data minimization approach favored by privacy scholars.

Access Requests. Interviewees were asked about subject access requests since the enforcement of the GDPR. While 11 companies stated that they had a standardized access request process (e. g., a website that presents the collected data to the user), slightly more (13) stated that they handled each SAR individually (e. g., the data is pulled from a database by the DPO). Responses show that overall fewer individuals than expected requested access. 12 participants stated that they received the expected amount and ten stated that they got "*less*" (4) or "*way less*" (6) than expected (one company preferred not to disclose). One interview partner (I-7) reported that they and other companies had expected *"a five-digit figure of requests per month"*, but over the whole year, they had received less than ten access requests. Consequently, half of the companies (12) do not consider changing how they handle access requests; only six of the participants reported that they were planning to change the process. Based on our data, there is no indication that companies that offer online tools receive more SARs than companies that only provide an offline tool. However, only one company stated that they monitor the number of requests they receive via their online platform. The remaining interviewees said they would like to change the process or reported that it had already been changed.

Regarding guidelines, 15 companies stated that they preferred more unified guidelines about how access requests should be handled. When asked who should provide this guidance, nine participants would prefer self-regulation (e. g., by the IAB), ten normative guidance (e. g., ISO standard), and three legislative guidance (e. g., amendments of the GDPR). These rather high numbers are likely related to the uncertainties that come with the new legislation. In the interviews especially representatives from smaller companies said they had a hard time defining processes they thought would comply with the new legislation. Thus, the desire for more regulation or guidelines could be a result of frustration companies faced when trying to comply with such complex regulations.

Challenges. In the online survey, we asked participants about the challenges for transparency and access requests. The answers to the questions can be grouped into two categories: (1) identification of users and (2) uncertainty about data flows. One of the main challenges for companies is to decide what level of identification is required before answering access requests. One the one hand, it is necessary to make sure that access is only granted to the actual individual whose data is stored. On the other hand, *"authentication of individuals creates more sensitive data"* (P-16).

The challenge of identification was also highlighted during the interviews. All but one of the participants agreed that identification is problematic since a cookie ID is not really sufficient to identify a person. One company stated that they were not concerned about cookie ID misuse, and all other companies implement the process differently, ranging from signed affidavits to screenshots of browser cookie stores. In terms of provided data, all companies claimed that they provide all data upon request—showing the clear discrepancy between user perception, our measurements, and company claims. The second challenge is related to the ad ecosystem that depends on data exchange between companies. To provide access to data, services have to *"understand the data flow of each partner [they] work with"* (P-13). While some companies most likely created a record of processing activities as required by the GDPR, one of the interview partners claimed that not all are aware of all data flows: *"[...] one of the reasons is technical, because I'm quite sure that they do not know how much data they have. I'm positive on this."* (I-1).

Opinions. In the last part of the survey as well as in the interviews, participants were asked about their personal opinion regarding the GDPR and new transparency rights. They assessed that the GDPR brought more attention to privacy in terms of management, mostly because of the high fines, but also because of public awareness. Therefore, all agreed that privacy has become more integrated into daily practices. If participants had an opinion on why users have such a bad view of ad-tech companies, they said that the main reason was the general negative attitude towards advertisements and/or data leaks, in general, was the main reason for that. One participant was puzzled by our finding that 65 % of users believe that companies do not share all data upon request and stated that from his/her point of view this was not correct.

Regarding transparency, all participants agreed that this was a positive development but added that the current systems provides only a few benefits to users and is quite impractical in practice. As I-3 put it, *"it's not useless, but users—and that's why they are not contacting us—that do not want to be tracked already have an app [to block tracking]"*, while other stated that SARs and privacy policies are "*sufficient*" (I-1) measures. At the end of the interviews, we asked what the interviewees thought would help to bring more transparency to the online ad industry. Three interview participants described a system that would give users high-level information about types of data collected and summarized similarly to credit scores, while still offering the technical data to those who are interested. Furthermore, two partners highlighted that transparency needs to be addressed when the data is collected or shared and that SARs do not help with that. Another idea was to add standardized 'traffic sign' like symbols to websites, ad banners, or cookie banners that easily describe which data is collected and for what purpose. When asked why the provided data is often formatted in non-usable ways, partners named different reasons like the data would "*become stultified*" (I-1) if raw data was formatted in a (standardized) way. One participant said that they could not do that because then they would become a data processor and not only a "*data collector*" (I-5). Another participant was surprised that not all companies provided inferred data like segments: "*Then you have been given an incomplete dataset because that's the only way that it actually works well for an ad company is to have a better understanding of what your preferences are. So I think then it's time for a complaint.*" (I-6).

Summary. Our study of the company perspective on data transparency did not reveal a consensus across topics. While some would

prefer more detailed guidelines, others think the current implementation is sufficient. Similarly, approaches to subject access requests vary and there appears to be little support for improvements as the main objective is legal compliance instead of fostering understanding at the users' end. While most participants think there is value in transparency mechanisms, there still seems to be a mismatch between user expectations and the understanding of personal data in the industry that has not considered users' data personal data for so long and does not always seem aware of the actual data flows.

6 DISCUSSION AND RECOMMENDATIONS

The companies we surveyed reported that the number of users asking for access is rather low. One factor for this is probably that many consumers are not (yet) aware of these new opportunities, granted by the online tools and new regulation. Still, some companies make it hard for users to learn who collects their data, so that even those who know about the new tools or their rights might have a hard time executing them. When getting access, users prefer receiving inferred information (e. g., interest segments) rather than technical data. In our user study, participants strongly expressed their wish that the provided data should be *"less technical"* (P-43) and in a *"easy to read visualization"* (P-324). But not all companies can provide such information because it depends on the business the company is doing. Furthermore, companies should provide both high-level information that users can easily understand and the underlying raw data so that users can use a tool to analyze the data according to their own needs. The problems with privacy policies have long been known, but still, it seems to be up to research to develop better tools, as companies do not focus on understanding but more on legal compliance. As I-5 stated:"*I wrote something like a hundred of privacy policies and all of them, I think, lack of transparency.* [...] *I think that I can defend myself in front of an authority, but I think a normal user, like a person that would read a policy, they will not understand fully what is happening.*".

Over 60 % of the companies that participated in our survey expressed the wish for more regulatory guidance on the design of access request processes. Our research suggests that consumers would appreciate simplified and unified ways to obtain transparency. Based on our results, we recommend providing a visual overview (e. g., a workflow diagram) that describes what happens with the collected data, where it comes from, and with whom it is shared. Further research in this area is needed to evaluate designs. These designs could be used as blueprints for companies when designing/improving their SAR processes or their transparency guidelines in general. An industry-wide standard would allow users to compare two services regarding their privacy impact. Those that want to be transparent about their practices should start by educating users about what they do with personal data *before* collecting and presenting it, first on a high level with the option of downloading raw data. Previous research has underlined this need for information literacy [6].

It is the public mistrust in the data sharing industry that fueled harsher regulations. To counter misconceptions, companies need to improve public understanding of their practices. It could help to provide information on what is *not* done with collected data (e. g., a company might not collect the users' location but users might still wonder why it is not shown). As we found that users struggle to identify the companies that might collect personal data, it would be helpful to add "Provided by X" information in every ad banner.

7 LIMITATIONS

In our transparency tool analysis, we analyzed the data provided by 22 ad companies—a subset of all advertisement related companies. Our study focuses on rather big companies (in terms of presence on websites), we did not analyze approaches of smaller companies. Still, we were able to identify different transparency approaches and our interviews with smaller companies do not hint that we missed methodological different approaches. The analyzed profiles do not include all data a company might collect (e. g., not all interest segments or all demographic information) and hence our classification might be false at some points. However, omitting the four cases in which we were unsure would not fundamentally change the overall results and findings.

Our user study is based on participants from the US only. We decided *not* to recruit EU residents since we would have had to provide the online questionnaire in various languages to avoid any bias because users do not take the questionnaire in their native language. Furthermore, we expect that US residents have similar needs when it comes to transparency in online advertisements. Our survey is based on a small subset of companies willing to participate in our study and telephone interviews. Therefore, the views they presented might not be fully representative of the industry as a whole. Still, we identified a diverse set of opinions and hope that future work can broaden the empirical basis of our results.

8 RELATED WORK

In the following, we discuss work closely related to ours and compare prior work to our approach that was not already mentioned in Section 3.1. Personal data is often conceptualized as an economic asset of a company [2]. Business models are created based on the collection and aggregation of personal data (e. g., [1, 20]) and also malicious attempts to collect such data have been studied [53].

Subject Access Requests. The SAR process, introduced by the GDPR, gives users the right to access their data collected by an online service. Urban et al. measure the SAR process in detail and show that the process is heterogeneous in terms of obstacles, timing, and success between different actors in the online advertising economy and can get quite elaborate [52]. Boniface et al. analyze the tension between authentication and security when users perform a SAR [7] and discuss measures used to identify users and discuss threats (e. g., denial of access) of too harsh measures. Most recently, two studies showed how subject access requests can be misused to get access to personal data companies stored about other individuals [10, 15]. Both works spoofed their identity and filed a SAR using the spoofed identity and found that companies sometime carelessly share data without verifying the users' identity carefully. Degeling et al. analyzed the adoption and effect of the GDPR regarding privacy policies and cookie notices [14].

User Transparency Tools. Leon et al. [33] evaluate the usability of several tools to limit OBA. The authors conduct a laboratory study in which participants for example use tools to opt-out of OBA and

show that all tools at the time had serious usability flaws which lead to the misuse and misunderstanding of such tools. Andreou et al. [4] presented an analysis of ad transparency tools provided by *Facebook*. In their work, they analyze the messages presented by the social media platform that explain why users see specific ads. They find that these messages are often incomplete and misleading. A web browser extension that gives users a more equitable choice with regards to ad blocking was presented by Parra-Arnau et al. [40]. The extension gives users fine-grained control over the ads they see and helps them understand how their browsing data is used. In a study with 40 participants, the authors evaluate the performance of their tool and show that re-targeting is the most common ad strategy.

Melicher et al. investigates the users' perspective on perceived benefits and risks of online tracking by conducting 35 user interviews [38]. They find that users, on the one hand, want more control over tracking but on the other hand, are unwilling to put effort into actually taking control. Schaub et al. evaluate three different tools that are used to block online tracking (e. g., *Ghostery*) regarding their effectiveness to inform users that they are being tracked [44]. One result of their study is that the tools do not manage to inform users about tracking and some users believe that the analyzed tools track them. Bashir et al. analyzed "ad preference managers", a special kind of transparency tool, that allows users to see and edit the segments companies have inferred about them [5]. In a user study, the authors analyze the correctness and compare the composition of such tools. They found that only 27 % of participants state that shown interests are relevant for them. In this study, *Google*, *Facebook* and *Twitter* provided such tools. Most recently, Utz et al. analyzed user interactions with cookie banners [55]. They found that current implementations of such banners often do not allow for meaningful consent and that only very few users interact with the banners.

9 CONCLUSION

The ad industry tries to provide more transparency about its practices and the data collected in different ways. We studied implementations of new transparency and data access possibilities in the online advertising industry. By analyzing different transparency approaches of ad-tech companies, we identified three conceptual types of data companies provide users, if they ask for access: (1) tracking data, (2) segment data, and (3) raw technical data. Our research shows that not all companies disclose the necessary information and that many do it in a way that is not user-friendly. The participants in our user study struggled to understand and interpret the personal data they received after they had asked for access, especially if confronted with low-level technical data. Most users rated the provided data to be helpful (> 50 %) while "segment data" was the most popular category. Furthermore, we found that a large proportion of users (65 %) do not trust that companies provide all collected data upon request. When it comes to the identification of companies that provide an standard ad banner, we found that only 24 % of users would correctly identify the ad network, while 46 % named the advertised products company as ad provider. We surveyed data protection officers in different companies active in the advertisement ecosystem to better understand their perspectives. Participants reported that there were technical hurdles rooted in the complexity of the ecosystem that make it hard to disclose exact information. Most companies in the interviews and almost half in the survey (42 %) stated that they receive less SARs than expected and 63 % of participating companies expressed their wish for more guidance when designing SAR processes. We also found that companies still primarily focus on compliance instead of transparency for users. Regulatory authorities and industry associations therefore need to develop clear guidelines and consistent consumer facing portals to improve the situation.

ACKNOWLEDGMENTS

This work was partially supported by the Ministry of Culture and Science of the State of North Rhine-Westphalia (MKW grants 005-1703-0021 "MEwM" and Research Training Group NERD.nrw). We would like to thank the anonymous reviewers for their valuable feedback, Christine Utz at Ruhr University Bochum for her efforts proofreading this work, and Yana Koval also at Ruhr University Bochum for transcribing the interviews. Furthermore, we would like to thank all participants of our user and company study and especially thank the interview partners for their valuable insights. Any findings, conclusions, opinions, or recommendations stated in this work are those of the authors and do not necessarily reflect the views of the participants of the conducted interviews/user studies, or the sponsors.

REFERENCES

[1] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. 2014. The Web Never Forgets: Persistent tracking mechanisms in the wild. In *Proceedings of the 21st ACM Conference on Computer and Communications Security (CCS'14)*. ACM Press, New York, New York, USA, 674–689.

[2] Alessandro Acquisti, Curtis R. Taylor, and Liad Wagman. 2015. The Economics of Privacy. *Journal of Economic Literature* 52 (2015), 64.

[3] Amazon. 2018. Amazon's Mechanical Turk. https://www.mturk.com/ Accessed: 2019-02-05.

[4] Athanasios Andreou, Giridhari Venkatadri, Oana Goga, Krishna P. Gummadi, Patrick Loiseau, and Alan Mislove. 2018. Investigating ad transparency mechanisms in social media: A case study of Facebook's explanations. In *Proceedings of the 2018 Symposium on Network and Distributed System Security (NDSS'18)*. Internet Society, San Diego, CA, 15.

[5] Muhammad Ahmad Bashir, Umar Farooq, Maryam Shahid, Muhammad Fareed Zaffar, and Christo Wilson. 2019. Quantity vs. Quality: Evaluating User Interest Profiles Using Ad Preference Managers. In *Proceedings of the 2019 Symposium on Network and Distributed System Security (NDSS'19)*. Internet Society, San Diego, CA, 15.

[6] Bettina Berendt. 2012. Data Mining for Information Literacy. In *Data Mining: Foundations and Intelligent Paradigms*, Dawn E. Holmes and Lakhmi C. Jain (Eds.). Springer-Verlag, Cham, 265–297.

[7] Coline Boniface, Imane Fouad, Nataliia Bielova, Cédric Lauradoux, and Cristiana Santos. 2019. Security Analysis of Subject Access Request Procedures How to authenticate data subjects safely when they request for their data. In *Annual Privacy Forum (APF'19)*. Springer-Verlag, Berlin, Heidelberg, 20.

[8] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. In *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*. American Psychological Association, Washington, DC, US, 57–71. https://doi.org/10.1037/13620-004

[9] Randolph E Bucklin and Catarina Sismeiro. 2003. A model of web site browsing behavior estimated on clickstream data. *Journal of marketing research* 40, 3 (2003), 249–267.

[10] Matteo Cagnazzo, Thorsten Holz, and Norbert Pohlmann. 2019. GDPiRated–Stealing Personal Information On- and Offline. In *Proceedings of the 2019 European Symposium on Research in Computer Security (ESORICS'19)*. Springer-Verlag, Cham, 21.

[11] S. Chiasson, Y. Abdelaziz, and F. Chanchary. 2018. Privacy Concerns Amidst OBA and the Need for Alternative Models. *IEEE Internet Computing* 22, 2 (2018),

52–61.

[12] CNN Business. 2019. Amazon reportedly employs thousands of people to listen to your Alexa conversations. https://edition.cnn.com/2019/04/11/tech/amazon-alexa-listening/index.html Accessed: 2019-02-05.

[13] CONSENT project. 2017. CONSENT Report Summary. https://cordis.europa.eu/result/rcn/140471_en.html Accessed: 2019-02-05.

[14] Martin Degeling, Christine Utz, Christoper Lentzsch, Henry Hosseini, Florian Schaub, and Thorsten Holz. 2019. We Value Your Privacy ... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy. In *Proceedings of the 2019 Symposium on Network and Distributed System Security (NDSS'19)*. Internet Society, San Diego, California, USA, 20.

[15] Mariano Di Martino, Pieter Robyns, Winnie Weyts, Peter Quax, Wim Lamotte Lamotte, and Ken Andries. 2019. Personal Information Leakage by Abusing the GDPR "Right of Access". In *Proceedings of the 15th Symposium on Usable Privacy and Security (SOUPS'19)*. ACM Press, New York, NY, 16.

[16] Digital Advertising Alliance. 2018. DAA Self-Regulatory Principles. https://digitaladvertisingalliance.org/principles Accessed: 2019-02-05.

[17] Digital Advertising Alliance. 2018. Your Ad Choices. https://optout.aboutads.info/?c=2&lang=EN Accessed: 2019-02-05.

[18] Claire Dolin, Ben Weinshel, Shawn Shan, Chang Min Hahn, Euirim Choi, Michelle L. Mazurek, and Blase Ur. 2018. Unpacking Perceptions of Data-Driven Inferences Underlying Online Targeting and Personalization. In *Proceedings of the 2018 Conference on Human Factors in Computing Systems (CHI'18)*. ACM Press, New York, NY, USA, 1–12.

[19] eMarketer. 2017. Ad Blocking in the US: eMarketer's Updated Estimates and Forecast for 2014–2018. https://www.emarketer.com/Report/Ad-Blocking-US-eMarketers-Updated-Estimates-Forecast-20142018/2002044 Accessed: 2019-02-05.

[20] Steven Englehardt and Arvind Narayanan. 2016. Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM Conference on Computer and Communications Security (CCS'16)*. ACM Press, New York, NY, USA, 1388–1401.

[21] IAB Europe. 2018. IAB Europe Transparency & Consent Framework Policies. http://www.iabeurope.eu/tcfdocuments/documents/legal/currenttcfpolicyFINAL.pdf Accessed: 2019-02-05.

[22] European Interactive Digital Advertising Alliance. 2018. Your Online Choices. http://www.youronlinechoices.com Accessed: 2019-02-05.

[23] Benjamin Fabian, Tatiana Ermakova, and Tino Lentz. 2017. Large-scale Readability Analysis of Privacy Policies. In *Proceedings of the 2017 Conference on Web Intelligence (WI'17)*. ACM Press, New York, NY, USA, 18–25.

[24] Uwe Flick. 2014. *The SAGE handbook of qualitative data analysis*. Sage Publications Ltd., Thousand Oaks, CA, USA.

[25] Samuel Grogan and Aleecia M. McDonald. 2016. Access Denied! Contrasting Data Access in the United States and Ireland. *Proceedings on Privacy Enhancing Technologies* 3, 23 (07 2016), 191–211.

[26] Mireille Hildebrandt. 2012. The Dawn of a Critical Transparency Right for the Profiling Era. *Digital Enlightenment Yearbook* 1 (2012), 41–56. http://www.medra.org/servlet/aliasResolver?alias=iospressISBN&isbn=978-1-61499-056-7&spage=41

[27] Inc. 2019. Google Is Absolutely Listening to Your Conversations, and It Confirms Why People Don't Trust Big Tech. https://www.inc.com/jason-aten/google-is-absolutely-listening-to-your-conversations-it-just-confirms-why-people-dont-trust-big-tech.html Accessed: 2019-02-05.

[28] Interactive Advertising Bureau. 2017. Internet Advertising Revenue Report. https://www.iab.com/wp-content/uploads/2018/05/IAB-2017-Full-Year-Internet-Advertising-Revenue-Report.REV2_.pdf Accessed: 2019-08-27.

[29] Interactive Advertising Bureau Europe. 2017. European Digital Advertising market has doubled in size in 5 years. https://www.iabeurope.eu/research-thought-leadership/resources/iab-europe-report-adex-benchmark-2017-report/ Accessed: 2019-08-27.

[30] Musa J Jafar and Amjad Abdullat. 2009. Exploratory analysis of the readability of information privacy statement of the primary social networks. *Journal of Business & Economics Research* 7, 12 (2009), 123–142.

[31] Carlos Jensen and Colin Potts. 2004. Privacy Policies As Decision-making Tools: An Evaluation of Online Privacy Notices. In *Proceedings of the 2004 Conference on Human Factors in Computing Systems (CHI'04)*. ACM Press, New York, NY, USA, 471–478.

[32] Daehyeok Kim, Sooel Son, and Vitaly Shmatikov. 2016. What Mobile Ads Know About Mobile Users. In *Proceedings of the 2016 Symposium on Network and Distributed System Security (NDSS'16)*. Internet Society, San Diego, CA, 14.

[33] Pedro Leon, Blase Ur, Richard Shay, Yang Wang, Rebecca Balebako, and Lorrie Cranor. 2012. Why Johnny Can't Opt out: A Usability Evaluation of Tools to Limit Online Behavioral Advertising. In *Proceedings of the 2012 Conference on Human Factors in Computing Systems (CHI'12)*. ACM, New York, NY, USA, 589–598.

[34] LimeSurvey Project Hamburg. 2012. LimeSurvey: An open source survey tool. https://www.limesurvey.org/ Accessed: 2019-02-05.

[35] Bin Liu, Anmol Sheth, Udi Weinsberg, Jaideep Chandrashekar, and Ramesh Govindan. 2013. AdReveal: Improving Transparency into Online Targeted Advertising. In *Proceedings of the 12th ACM SIGCOMM Workshop on Hot Topics in Networks (HotNets'12)*. ACM Press, New York, NY, USA, 1–7.

[36] Miguel Malheiros, Charlene Jennett, Snehalee Patel, Sacha Brostoff, and Martina Angela Sasse. 2012. Too Close for Comfort: A Study of the Effectiveness and Acceptability of Rich-media Personalized Advertising. In *Proceedings of the 2012 Conference on Human Factors in Computing Systems (CHI'12)*. ACM Press, New York, NY, USA, 579–588.

[37] Aleecia M. McDonald and Lorrie Faith Cranor. 2010. Americans' Attitudes About Internet Behavioral Advertising Practices. In *Proceedings of the 9th ACM Workshop on Privacy in the Electronic Society (WPES '10)*. ACM Press, New York, NY, USA, 63–72.

[38] William Melicher, Mahmood Sharif, Joshua Tan, Lujo Bauer, Mihai Christodorescu, and Pedro Giovanni Leon. 2016. (Do Not) Track Me Sometimes: UsersâĂŹ Contextual Preferences for Web Tracking. *Proceedings on Privacy Enhancing Technologies* 2016, 2 (2016), 135 – 154. https://content.sciendo.com/view/journals/popets/2016/2/article-p135.xml

[39] Meik Michalke. 2018. koRpus: An R Package for Text Analysis. https://reaktanz.de/?c=hacking&s=koRpus Version: 0.11-5. Accessed: 2019-02-05.

[40] Javier Parra-Arnau, Jagdish Prasad Achara, and Claude Castelluccia. 2017. MyAdChoices: Bringing Transparency and Control to Online Advertising. *ACM Transactions on the Web* 11, 1, Article 7 (2017), 47 pages.

[41] Angelisa C. Plane, Elissa M. Redmiles, Michelle L. Mazurek, and Michael Carl Tschantz. 2017. Exploring User Perceptions of Discrimination in Online Targeted Advertising. In *Proceedings of the 26th USENIX Security Symposium (USENIX Sec'17)*. USENIX Association, Vancouver, BC, 935–951.

[42] Ashwini Rao, Florian Schaub, and Norman M. Sadeh. 2015. What do they know about me? Contents and Concerns of Online Behavioral Profiles. *CoRR* abs/1506.01675 (2015), 14. arXiv:1506.01675 http://arxiv.org/abs/1506.01675

[43] Sonam Samat, Alessandro Acquisti, and Linda Babcock. 2017. Raise the Curtains: The Effect of Awareness About Targeting on Consumer Attitudes and Purchase Intentions. In *Proceedings of the 13th Symposium on Usable Privacy and Security (SOUPS'17)*. ACM Press, New York, NY, 299–319.

[44] Florian Schaub, Aditya Marella, Pranshu Kalvani, Blase Ur, Chao Pan, Emily Forney, and Lorrie Faith Cranor. 2016. Watching Them Watching Me: Browser Extensions Impact on User Privacy Awareness and Concern. In *Proceedings of the 2016 Workshop on Usable Security (USEC'16)*. Internet Society, Reston, VA, 10.

[45] Oleksii Starov and Nick Nikiforakis. 2017. Extended Tracking Powers. In *Proceedings of the 26th World Wide Web Conference (WWW '17)*. ACM Press, New York, New York, USA, 1481–1490.

[46] State of California. 2018. California Consumer Privacy Act of 2018. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375 Accessed: 2019-02-05.

[47] The European Parliament and the Council of the European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union, L 119/1.

[48] The Guardian. 2018. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election Accessed: 2019-02-05.

[49] The Verge. 2019. Facebook also hired human contractors to listen to audio from its Messenger app. https://www.theverge.com/2019/8/13/20804315/facebook-messenger-audio-conversations-listening-human-contractors Accessed: 2019-02-05.

[50] Triple Lift. 2018. GDPR–Your Individual Rights. https://access.triplelift.com/ Accessed: 2019-02-05.

[51] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. 2012. Smart, Useful, Scary, Creepy: Perceptions of Online Behavioral Advertising. In *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS'12)*. ACM Press, New York, NY, 1–15.

[52] Tobias Urban, Dennis Tatang, Martin Degeling, Thorsten Holz, and Norbert Pohlmann. 2019. A Study on Subject Data Access in Online Advertising after the GDPR. In *Proceedings of the 2019 International Workshop on Data Privacy Management (DPM'19)*. Springer-Verlag, Cham, 18.

[53] Tobias Urban, Dennis Tatang, Thorsten Holz, and Norbert Pohlmann. 2018. Towards Understanding Privacy Implications of Adware and Potentially Unwanted Programs. In *Proceedings of the 2018 European Symposium on Research in Computer Security (ESORICS'18)*. Springer-Verlag, Cham, 449–469.

[54] U.S. Census Bureau. 2017. 2017 American Community Survey. https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t Accessed: 2019-02-05.

[55] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. 2019. (Un)informed Consent: Studying GDPR Consent Notices in the Field. In *Proceedings of the 2019 ACM Conference on Computer and Communications Security (CCS'19)*. ACM Press, New York, NY, USA, 19.

[56] Narseo Vallina-Rodriguez, Srikanth Sundaresan, Abbas Razaghpanah, Rishab Nithyanand, Mark Allman, Christian Kreibich, and Phillipa Gill. 2018. Apps, Trackers, Privacy, and Regulators: A Global Study of the Mobile Tracking Ecosystem. In *Proceedings of the 2018 Symposium on Network and Distributed System Security (NDSS'18)*. Internet Society, San Diego, CA, 15.
[57] Giridhari Venkatadri, Alan Mislove, and Krishna P. Gummadi. 2018. Treads: Transparency-Enhancing Ads. In *Proceedings of the 17th ACM SIGCOMM Workshop on Hot Topics in Networks (HotNets'18)*. ACM Press, New York, NY, USA, 169–175.
[58] Shuai Yuan, Jun Wang, and Xiaoxue Zhao. 2013. Real-time Bidding for Online Advertising: Measurement and Analysis. In *Proceedings of the Seventh Workshop on Data Mining for Online Advertising (ADKDD'13)*. ACM Press, New York, NY, USA, 1–8.

A USER SURVEY QUESTIONS

In our user study, we used the following questionnaire to evaluate user perception of current transparency tools provided by different ad-tech comapnies.

All questions, excluding the consent form and open-ended questions (which can be left blank), offer an *"I prefer not to answer"* option we omitted in the following for readability and space-saving. If not stated otherwise, we used: (5PL) 5 point Likert scales, (Y/N) yes, no, I do not know, or (OE) open-ended answer options (AO).

Question Group: **Self Assessment**
Q1: *Online advertisements or recommendations (e.g., for product items or articles) I see suit my interests.* AO: 5PL from "Always" to "Never".
Q2: *Rate your personal experience with ads or recommendations (e.g., for product items or articles) you see online.* AO: 5PL from "Very satisfied" to "Very dissatisfied".
Q3: *Have you ever wondered why you see a specific ad or recommendation (e.g., for product items or articles) online?* AO: 5PL from "Always" to "Never".
Q4: *Have you ever requested a copy of the personal data that a company has collected about you?* AO: Y/N
Q5: *I think having access to data ad companies collected about me is useful to better understand how they use my personal data.* AO: 5PL from "Strongly Agree" to "Strongly Disagree".
Q6: *How do you rate your knowledge about online advertisement?* AO: 5PL from "Not knowledgeable" to "Very knowledgeable".
Q7: *Which of the following statements regarding the online advertisements do you think are true?* AO (multiple choice): (1) "Ad companies might share my personal data with their partners", (2) "Ad companies have access to my full browsing history", (3) "Ad companies collect personal data about me whenever they show me an ad", (4) "Ad companies know which device I am using", (5) "Ad companies have access to all products I bought online", (6) "Ad companies learn my interests from my online actions", and (7) "All statements are false"
Q8: *Do you use any of the following mechanisms to protect your privacy online?* AO (multiple choice): (1) "I use a browser extension to block ads or tracking (e. g., Ghostery, AdBlock Plus, Privacy Badger, Disconnet)", (2) "I browse in private mode ("incognito mode") or use a VPN from time to time", (3) "I delete cookies from my browser from time to time", (4) "I opted out of online behavioral advertising with at least one company", and (5) "None of the above".
Q9: *Do you have any comments regarding your experience with online advertisements?* AO: Open-ended
Question Group: **Identifying Responsible Parties**
Q10: *Take a look at the highlighted part of the following picture (red frame)[Figure 1a]. If you wanted to understand on which data the specific ad or recommendation is based, whom would you ask?* AO (multiple choice): (1) "The Review Experts", (2) "ESPN.com", (3) "Mansion Global", (4) "ZaloTech", (5) "Outbrain", (6) "I do not know".
Q11: *Take a look at the highlighted part of the following picture (red frame)[Figure 1b]. If you wanted to understand on which data the specific ad or recommendation is based, whom would you ask?* AO (multiple choice): (1) "Reddit.com", (2) "The Outnet.com", (3) "Google", and (4) "I do not know".
Question Group: **Transparency Tool Assessment**
On the next pages you will see three different ways how companies make personal data accessible that they collected about someone. The categories are: Technical data. Information automatically transmitted when you surf the web. Tracking data. Information on which websites the company tracked you. Interest data. User interests the company interfered from the collected data. We will provide you profiles of three different companies in each category (nine in total) that were collected about the same individual. Note: Companies may provide different information as they have different data sources
Q12-Q21: *Note: The following four questions were asked to each profile displayed in Figures 3, 4, and 5 (i. e., each question was asked 9 times). AO: (5PL) from "Strongly Agree" to "Strongly Disagree".* (1) "This is the kind of information I expected to see.", (2)"The website displays helpful information regarding personal data collected about me.", (3) "I understand the presented information.", and (4) "The information is presented in a clear way."
Question Group: **General Transparency Questions**
Q22: *In general, do you think that companies provide all personal data they actually collected about you if you request them?* AO: Y/N.
Q23: *Prioritize which of the following information should be included if you request access to your personal data. By technical (raw) data we mean data displayed in the following images: [Figure 4] By tracking data we mean data displayed in the following images: [Figure 5] By interest data we mean data displayed in the following images: [Figure 3] You can skip this question if you prefer not to answer* AO: Ranking from 1 to 3 for each profile.
Q24: *In this survey you saw personal data ad companies collected about a stranger, are you now interested in personal data collected about you?* AO: 5PL from "Very interested" to "Not at all interested".
Q25: *Knowing what data ad companies collect about me, I would reconsider my online behaviour.* AO: 5PL from "Strongly Agree" to "Strongly Disagree".

Question Group: **Improvement Suggestions**
Q26:*From your point of view what can ad companies do better to increase transparency in the online ad ecosystem?* AO: open-ended
Q27: *How can ad companies improve the presentation of collected personal data?* AO: open-ended
Question Group: **Demographics**
Q28: *How old are you?* AO: (1) "18-24", (2) "25-34", (3) "35-44", (4) "45-54", (5) "55-65", and (6) "65 years or older".
Q29: *How do you identify?* AO: (1) "Female", (2) "Male", and (4) "Non-binary"
Q30: *What is the highest degree or level of school you have completed? (If you're currently enrolled in school, please indicate the highest degree you have received.)* AO: (1) "Less than a high school diploma", (2) "High school graduate", (3) "Bachelor's degree", (4) "Master's degree", (5) "Professional degree", and (6) "Doctorate degree".

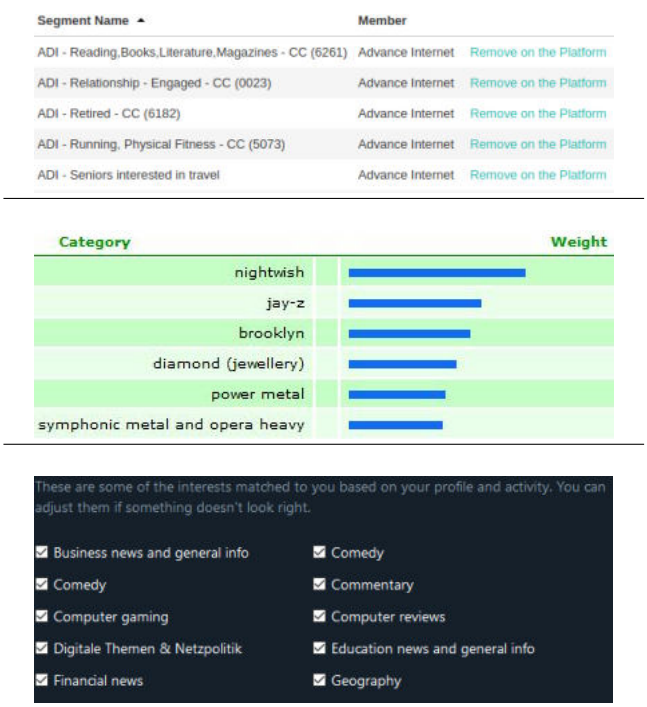


Figure 3: Inferred interest segments provided by transparency tools to users (anonymized for submission)

B COMPANY SURVEY QUESTIONS

We used the following questionnaire to identify problems companies faced when implementing transparency tools.

All questions, excluding the consent form and open-ended questions (which can be left blank), offer an *"I prefer not to answer"* answer option which is omitted in the following for readability and space-saving.

Question Group: **General Questions**
Q1: *What is your role in your company?* AO: (1)Legal / Privacy Management, (2) General Support / Helpdesk, (3) Data Protection Officer (DPO), (4) External / Consulting, and (5) Other(s) (please specify).
Q2: *Who is responsible to handle Subject Access Requests in your company?* AO: (1)Legal / Privacy Management, (2)General Support / Helpdesk, (3) Data Protection Officer (DPO), (4)External / Consulting, and (5) Other(s) (please specify).
Q3: (optional) *How high is the percentile amount of your individual data subjects that actually perform a Subject Access Request (e.g.,"1 out of 10.000 data subjects" or "1%")?* AO: open-ended
Q4: *Considering your expectations before the GDPR took effect: How often do you handle Subject Access Requests in your company?* AO: 5PL from "Way more than expected." to "Way less than expected. "
Question Group: **Development of the SAR process**
Q5: *Do you have a standardized process to handle Subject Access Requests in your company?* AO: (1) Yes, there is a (semi)automated process, (2) No, each request is answered individually, or (3) I do not know.
Q6: *Now, six months after the GDPR took effect, do you think it is necessary to change the way you handle Subject Access Requests?* AO: (1) We already changed the way we handle such requests (since the GDPR took effect), (2) Yes, we plan to change the way we handle such requests, (3) No, but I think we should change the way we handle such requests, or (4) No.
Q7: *Should there be a detailed guideline on how to handle Subject Access Requests? If so, who should provide it?* AO: (multiple-choice) (1) Industry self-regulation (e. g., IBA or DAA), (2) Normative regulation (e. g., in a standard/norm provided by ISO), (3) Legislative regulation (e.g., as amendment of the GDPR), or (4) No more regulation is needed.
Q8: (optional)*What do you think were the biggest obstacles when designing your Subject Access Request process?* AO: open-ended
Question Group: **Views on transparency**
Q9: (optional) *Which benefits does the GDPR provide for your company?* AO: open-ended
Q10:(optional) *To what extend are Subject Access Requests a useful tool for users to regain control of their data?* AO: open-ended
Q11: (optional) *Do you think that the GDPR provides any benefits to users when it comes to transparency in the online advertising industry?* AO: open-ended
Question Group: **Demographics**
Q12: *Is the headquarter of your company located in a country that is part of the European Union?* AO: (1) Yes, (2)No (please specify), or (3) I do not know.
Q13: *In which segments of the digital advertising ecosystem is your company active?* AO (multiple choice): (1) Agency / Agency Trading Desk, (2) Targeting / Audience, (3) Data

Manament Platform / Data Provider, (4) Ad Exchange / Ad Server, (5) Verification & Privacy, (6) Demand Side Platform, (7) Ad network / Supply Side Platform, or (8) Other(s) (please specify).
Q14: *How many people are employed at your company?* AO: (1) 1–50, (2) 51–100, (3) 101–250, (4) 251–500, (5)501–1,000, or (6)> 1,000.
Q15: *How many employees work in your department?* AO. (1) 1, (2) 2–5, (3) 6–10, (4) 11–20, or (5) > 20.


```
{
  "cookieId" : " DR-000000043O8871417",
  "ipAddresses" : [ "161.185.160.XXX" ],
  "locations" : [ {
    "city" : "hoboken",
    "state" : "new jersey",
    "country" : "united states",
    "postal" : "07030"
  } ]
}
```


timestamp	Dec. IP address	Internet Service Provider	Continent	Numeric City ID	Postal code	Time zone	Browser ID	OS ID
6/19/18 4:29	-1972356321	Spectrum	us	7540	7030	-500	30	43
6/19/18 4:32	-1972356321	Spectrum	us	7540	7030	-500	30	43
6/19/18 4:32	-1972356321	Spectrum	us	7540	7030	-500	30	43
6/19/18 4:32	-1972356321	Spectrum	us	7540	7030	-500	30	43
6/19/18 4:32	-1972356321	Spectrum	us	7540	7030	-500	30	43
6/19/18 4:32	-1972356321	Spectrum	us	7540	7030	-500	30	43
6/19/18 4:32	-1972356321	Spectrum	us	7540	7030	-500	30	43
6/19/18 4:37	-1972356321	Spectrum	us	7540	7030	-500	30	43
6/19/18 4:52	-1972356352	Spectrum	us	7540	7030	-500	30	43
6/19/18 4:52	-1972356321	Spectrum	us	7540	7030	-500	30	43
6/19/18 4:52	-1972356321	Spectrum	us	7540	7030	-500	30	43
6/19/18 4:52	-1972356321	Spectrum	us	7540	7030	-500	30	43
6/19/18 4:52	-1972356352	Spectrum	us	7540	7030	-500	30	43

time	u_ip	ua	kvClob
24.09.18 09:29:25 UTC	127.0.0.0	Mozilla/5.0 (X11; Ubuntu	cs=-;fst=-;c=<Country>;st=<state>;
24.09.18 09:28:04 UTC	127.0.0.0	Mozilla/5.0 (X11; Ubuntu	cs=-;fst=-;ip=127.0.0.0/24;hb=1;de
24.09.18 09:24:19 UTC	127.0.0.0	Mozilla/5.0 (X11; Ubuntu	cs=-;fst=-;c=<Country>;st=<state>;
24.09.18 09:24:18 UTC		Mozilla/5.0 (X11; Ubuntu	c=DE;st=<state>;dma=<metroKey>
24.09.18 09:17:29 UTC	127.0.0.0	Mozilla/5.0 (X11; Ubuntu	cs=-;fst=-;c=<Country>;st=<state>;
24.09.18 09:16:19 UTC	127.0.0.0	Mozilla/5.0 (X11; Ubuntu	cs=-;fst=-;c=<Country>;st=<state>;
24.09.18 09:16:17 UTC	127.0.0.0	Mozilla/5.0 (X11; Ubuntu	cs=-;fst=-;c=<Country>;st=<state>;
24.09.18 09:10:44 UTC	127.0.0.0	Mozilla/5.0 (X11; Ubuntu	cs=-;fst=-;c=<Country>;st=<state>;

Figure 4: Technical data provided by transparency tools to users (anonymized for submission).

Last seen on	Site	IP
2018-09-24 08:51:01	https://www.01net.com/	127.0.0.0
2018-09-24 08:53:11	https://ad3.adserver01.de/www/delivery/af	127.0.0.0
2018-09-24 09:05:06	https://www.farfeshplus.com/	127.0.0.0
2018-09-24 09:15:50	https://www.wetter.de/	127.0.0.0
2018-09-24 09:16:31	https://www.pistonheads.com/	127.0.0.0
2018-09-24 09:18:59	https://ad6.ad-srv.net/request_content.php?	127.0.0.0

time	url	ip	gdprQCConsent	cookieIn	type
Mon Sep 24 09:29:25	https://www.ladepeche.fr/	-1033994496	1	5ba8a46b-987b9-2ce27-93e48	APPNEXUS
Mon Sep 24 09:28:04	https://www.ebay-kleinanzeigen.de/	-1033994496	1	5ba8a46b-987b9-2ce27-93e48	CASALE
Mon Sep 24 09:24:19	http://dailycaller.com/section/daily-vaper/	-1033994496	1	5ba8a46b-987b9-2ce27-93e48	APPNEXUS
Mon Sep 24 09:24:18	http://segapi.quantserve.com/api/segments.json?	-1033994496	1		
Mon Sep 24 09:17:29	http://style24.it	-1033994496	x	5ba8a46b-987b9-2ce27-93e48	BIDSWITCH
Mon Sep 24 09:16:19	https://www.pistonheads.com	-1033994496	1	5ba8a46b-987b9-2ce27-93e48	GOOGLE
Mon Sep 24 09:16:17	https://www.pistonheads.com/	-1033994496	1	5ba8a46b-987b9-2ce27-93e48	APPNEXUS
Mon Sep 24 09:10:44	https://myfav.life/	-1033994496	1	5ba8a46b-987b9-2ce27-93e48	APPNEXUS
Mon Sep 24 09:09:36	https://www.kompas.com/	-1033994496	1	5ba8a46b-987b9-2ce27-93e48	PUBMATIC
Mon Sep 24 09:05:05	https://www.farfeshplus.com	-1033994496	1	5ba8a46b-987b9-2ce27-93e48	GOOGLE
Mon Sep 24 08:58:43	https://www.journaldesfemmes.com/	-1033994496	1	5ba8a46b-987b9-2ce27-93e48	APPNEXUS

Buyer Member ID	Buyer Member Name	UID
2636	PulsePoint DSP	5gToVhdcg2bk
3335	AppNexus DSP	426514054948993099

Figure 5: Tracking data provided by transparency tools to users (anonymized for submission).