Cookie Banners and Privacy Policies: Measuring the Impact of the GDPR on the Web

MICHAEL KRETSCHMER, RWTH Aachen University, Germany
JAN PENNEKAMP* and KLAUS WEHRLE, RWTH Aachen University, Germany

The General Data Protection Regulation (GDPR) is in effect since May of 2018. As one of the most comprehensive pieces of legislation concerning privacy, it sparked a lot of discussion on the effect it would have on users and providers of online services in particular, due to the large amount of personal data processed in this context. Almost three years later, we are interested in revisiting this question to summarize the impact this new regulation has had on actors in the World Wide Web. Using Scopus, we obtain a vast corpus of academic work to survey studies related to changes on websites since and around the time, the GDPR went into force. Our findings show that the emphasis on privacy increased w.r.t. online services, but plenty potential for improvements remains. Although online services are on average more transparent regarding data processing practices in their public data policies, a majority of these policies still either lack information required by the GDPR (e.g., contact information for users to file privacy inquiries), or do not provide this information in a user-friendly form. Additionally, we summarize that online services more often provide means for their users to opt out of data processing, but regularly obstruct convenient access to such means through unnecessarily complex and sometimes illegitimate interface design. Our survey further details that this situation contradicts the preferences expressed by users both verbally and through their actions, and researchers have proposed multiple approaches to facilitate GDPR-conform data processing without negatively impacting the user experience. Thus, we compiled reoccurring points of criticism by privacy researchers and data protection authorities into a list of four guidelines for service providers to consider.

CCS Concepts: • **Security and privacy → Web application security**; *Privacy protections*; *Usability in security and privacy*;

Additional Key Words and Phrases: Cookies; Privacy; GDPR; Web; Privacy Legislation; Fingerprinting

ACM Reference Format:
Michael Kretschmer, Jan Pennekamp, and Klaus Wehrle. 2021. Cookie Banners and Privacy Policies: Measuring the Impact of the GDPR on the Web. *ACM Trans. Web* 15, 4, Article 20 (June 2021), 42 pages. https://doi.org/10.1145/3466722

1 INTRODUCTION

In recent years, a number of high-profile instances relating to the processing of personal information from web services, such as the publication of the US National Security Agency program Prism [61] and the Cambridge Analytica scandal [142], have repeatedly sparked discussions about online privacy. As a result, people distrust online services and especially social networks [65] that collect

*Corresponding author: Tel: +49-241-80-21411, Fax: +49-241-80-22222

Authors' addresses: Michael Kretschmer, michael.kretschmer@rwth-aachen.de, RWTH Aachen University, Ahornstr. 55, Aachen, Germany, 52074; Jan Pennekamp, jan.pennekamp@comsys.rwth-aachen.de; Klaus Wehrle, klaus.wehrle@comsys.rwth-aachen.de, Communication and Distributed Systems, RWTH Aachen University, Ahornstr. 55, Aachen, Germany, 52074.

1559-1131/2021/6-ART20 $15.00
https://doi.org/10.1145/3466722

data directly related to their person. Although the EU-wide *Data Protection Directive* [48], which was intended to protect individual privacy, was passed already in 1995, it is unsuited to effectively accomplish this task in today's online ecosystem [38]. This failure is partly due to the drastically increased capability to collect, store, and process data [6], but also due to weak enforcement and a lack of fines for offenders [70]. The most recent piece of broad EU-privacy legislation is the *General Data Protection Regulation* (GDPR) [51], which introduces binding rules concerning privacy to all EU states for the first time and is in effect since May 2018.

In a report from 2016, Englehardt and Narayanan [42] showed that out of the worldwide one million most popular websites at least 70 % employ some form of tracking, which is affected by the GDPR. For US American websites, this number may be as high as 92 % [99]. The reception of this regulation is highly heterogeneous ranging from arguments that it places an almost impossible burden on providers of online services [66], over its wording being too ambiguous to make any prediction on the effect it will have [70, 84], to it changing the world for the better [4]. Since the GDPR's enactment, numerous studies [31, 74, 93, 94, 118, 119, 129] have analyzed data collected from the web to investigate the consequences of the GDPR. Although this work is fundamental to informed discussions about the impact of the GDPR, the highly specific nature of such work also limits its scope. Due to the number of novel requirements for service providers and rights granted to their users by the GDPR, we identify the need for an extensive literature study as it is more suited to summarize the far-reaching consequences the GDPR has had on the World Wide Web.

Like previous surveys, e.g., Laperdrix et al. [90], we intend to provide interested readers a detailed overview into the state of the art w.r.t. the implications of the GDPR on the web. As such, we do not aspire to cover the entirety of one specific form of tracking as pursued by Laperdrix et al. [90], but instead consider a wide range of aspects related to online tracking with special emphasis on cookies, cookie consent notices (cookie banners), privacy policies, and browser fingerprinting, and how they were affected by the GDPR. As a foundation for our literature survey, we rely on Scopus[1], a popular literature database. Consequently, we compile and analyze an exhaustive corpus in the scope of tracking on the web published around the enactment of the GDPR and beyond.

Contributions. Three years after the GDPR has been enacted, we aim to facilitate a nuanced and objective discussion in this paper about its impact on the web, relying on factual findings [31, 93, 94, 119, 149, 157] as well as academic discussions from various points of view [29, 38, 57, 66, 109]. Our main contributions in this work are as follows.

(1) We establish a broad background of essential information regarding privacy legislation and tracking on the web based on the results of our literature study. This knowledge helps the reader to interpret the empirical results we summarize in this work, and serves as an introduction point to the vast (and captivating) topic of online privacy.
(2) We present an extensive summary of key findings based on scientific research regarding the impact of the GDPR. We learn that although the GDPR can directly be linked to an increase in cookie consent notices close to 40 % and the volume of privacy policies of around 60 %, there is no strong evidence that the GDPR has led to a decrease in online tracking.
(3) We outline specific suggestions of metrics related to online privacy that warrant further investigation. We identify the need to work towards standardizing measurement methodologies, especially in terms of tracking analysis, to allow the comparisons of study results, which could fill in some blanks regarding the GDPR's impact on online privacy and the evolution of tracking technology.
(4) We point out the most pervasive challenges for service providers to reach compliance with the GDPR, based on our research: We find that insufficient control is commonly provided to

[1]https://www.scopus.com/ — "Expertly curated abstract & citation database" for academic articles.

the users of online services by making it either difficult or impossible for them to opt out of non-essential data processing and to access or delete their personal data. Furthermore, we can attribute these shortcomings in many cases to the underlying business model of the services, as well as a lack of involvement by the responsible data protection authorities.

(5) We contrast the arguments brought up in scientific and journalistic publications on GDPR-related controversial topics. Thereby, we present the differing points of view on the not yet fully understood implications of the GDPR, such as the right to data portability, the use of Blockchain technology, and privacy by default.

To the best of our knowledge, we provide the first extensive literature review on the impact of the GDPR on the web.

Paper Organization. First, we briefly describe the methodology of our literature survey along with the target audience of this work in Section 2. Then, in Section 3, we provide a comprehensive summary of EU privacy legislation, as well as the different technologies used to track individuals on the web. Afterward, in Section 4, we detail the findings of GDPR-focused related work along with commonly described challenges in obtaining these results. In Section 5, we discuss the limitations both regarding the collection of empirical data relevant to our research and the GDPR itself, followed by an outlook on future developments related to provisions of the GDPR, such as data portability and privacy by default. Subsequently, in Section 6, we derive recommendations for service providers based on the findings that we presented and analyzed in previous sections. Finally, Section 7 concludes this paper and includes a future work outlook.

2 SURVEY METHODOLOGY & AUDIENCE OF OUR WORK

To create a well-founded foundation for our literature survey, we define a specific focus for our work. We mainly focus on scientific publications to gather quantitative data regarding the GDPR's impact. Due to its significant coverage, we decided to rely on the literature database *Scopus*. We searched for all relevant scientific literature by querying for all English publications between 2018 and February 2021 containing the phrases *GDPR* and either *cookies*, *fingerprinting*, *tracking*, or *web* in title, abstract, or keywords. In total, Scopus returned 172 resources, of which we immediately excluded 21 because they constituted complete proceedings and one paper which appeared twice.

For the remaining papers, we read the abstract of every paper and skimmed through the introduction and result section if we were unable to form a well-founded opinion solely based on the abstract whether the paper fits into the scope of our analysis. Through this process, we additionally eliminated 45 papers because they had no noticeable relation to the web. We further excluded 56 publications for not containing any measurement results. Most of these non-quantitative studies introduced novel techniques related to data management and privacy with a special emphasis on approaches utilizing or extending the Layered Privacy Language, which Gerl et al. [62] introduced in 2018. The most prominent research areas in these excluded papers were related to medicine and the Internet of Things.

We augmented this initial set of papers with selected publications that were not indexed in Scopus by following the references within our corpus. Thus, we obtained an extensive collection, which describes the state of the art. To present a diverse set of viewpoints, we also consider non-academic references. In particular, we further rely on factual journalistic reports [7, 14, 21, 72] to highlight notable events related to the GDPR's impact on the web. Additionally, we include some more opinionated online resources [28, 125, 135] in an effort to contrast contrary stances on the GDPR and informational material from reputable (mostly governmental) online sources [10, 33, 84, 86].

With this work, we address a broad audience that is interested in the topic of privacy on the web and what effect recent privacy legislation, with a particular focus on the GDPR, has had. Our goal is to highlight persisting privacy threats for online users and discuss their origin based on

the findings of academic research. We want to address readers irrespective of their background in privacy and the GDPR by providing a concise yet well-founded overview in this recently emerging area (the GDPR has a significantly larger scope than previous privacy legislation). At this point, we want to stress that this work does *not* (primarily) intend to define a roadmap by meticulously outlining all open research gaps for future scientific research in this domain. Instead, we mainly focus on our survey of existing work to look into the impact the GDPR has (had) on the web. Likewise, we cannot provide warranted legal advice and suitable concrete configuration guidelines for web service administrators as we have no profound background in law. Nevertheless, this work references all relevant legislation for interested readers.

To summarize, this work is meant to serve as an introductory work for a readers with varying background by detailing technical and legal developments, as well as open challenges as reported by related work.

3 BACKGROUND

The GDPR is often cited as the most comprehensive piece of legislation regarding privacy [66, 70, 94]. As such, its enactment impacted a wide variety of common practices implemented by a majority of contemporary applications and software systems. Thus, we introduce the legal background of personal data processing within the European Union and how it changed over the years leading up to the GDPR in Section 3.1. Subsequently, we summarize the GDPR in Section 3.2, primarily focusing on GDPR articles that were discussed in the work we analyzed. Finally, in Section 3.3, we present the most common tracking techniques as well as measures service providers have adopted to increase transparency and privacy in the interest of their users. In the remainder of this paper, we then consider both areas, legislation and technical building blocks, jointly.

3.1 Pre-GDPR Privacy Legislation in the European Union

We first provide an overview of the history of European privacy legislation. As the number of directives and regulations related to this topic is vast, we focus on an overview of the most comprehensive and relevant changes in legislation.

Data Protection Directive. The Data Protection Directive [48], passed in 1995, is considered to be the first EU-wide binding legislation regulating the processing of data relating to natural persons. It laid out basic definitions still relevant to today's privacy legislation, e.g., defining personal data as "any information relating to an identified or identifiable natural person" [48, Article 2(a)] along with what qualifies as consent [48, Article 2(h)]. The individual, which this personal data concerns, is, in this case, referred to as *data subject*, and the processing party is called *data controller*. This directive does not only apply to processing that takes place in member states of the EU, but to every processing that involves personal data of citizens of any EU state [48, Article 25]. However, directives differ from regulations as they set specific goals that are supposed to be achieved through national laws, instead of enacting such laws directly [52]. Thus, the effect of directives may be different for EU citizens depending on the member state they inhabit.

Privacy and Electronic Communications Directive. In 2002, the Privacy and Electronic Communications Directive [49], also known as ePrivacy Directive, was passed. It builds on the Data Protection Directive and refines formulations concerning Internet data traffic. The introduction of the concept of *informed consent* is most relevant to this work. Informed consent mandates services to inform end-users what kind of data is stored on their device, most notably cookies, and for what specific purpose. Furthermore, every end-user must be granted the opportunity to refuse such storage, provided that it is not strictly needed for a (regular) operation of the respective service. This directive was amended in 2009 [50] to address technological advances that mostly concern mobility. Furthermore, it replaced the previous article on informed consent with the concept of *explicit*

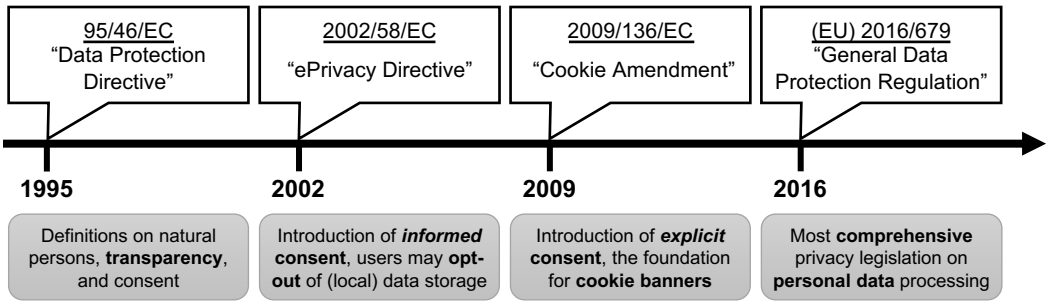


Fig. 1. Timeline of EU privacy legislation featuring four pieces of legislation that are most relevant to our work. These examples demonstrate the continuous adaptations that are necessary to limit privacy risks stemming from, among other factors, a significant increase in the capability to collect personal data.

consent, i.e., end-users must explicitly agree before any information can be stored on their devices. This amendment also allows end-users to withdraw consent relating to the storage and processing of their personal data at any time. Figure 1 illustrates how privacy legislation got progressively more comprehensive over time in the EU.

Takeaway. *The Data Protection Directive and ePrivacy Directive are the central pieces of privacy legislation passed in the EU prior to the GDPR. The ePrivacy Directive requires data controllers (party processing data) to obtain the explicit consent of data subjects (individuals whose data is affected) before processing their personal data. However, the effectiveness of these directives is impeded by lacking a binding set of binding fines for misbehaving entities and have contributed to differing privacy standards throughout the EU.*

3.2 The General Data Protection Regulation (GDPR)

Building upon the existing legislation, the European Parliament passed the General Data Protection Regulation in 2016. The GDPR establishes new restrictions on data processing, obligations for data controllers, and freedoms for data subjects, tailored to address current privacy concerns [135]. In the following, we first give an overview of the GDPR (Section 3.2.1) before focusing on specific articles, especially relevant to our research (Section 3.2.2) and briefly addressing the planned *ePrivacy Regulation* (Section 3.2.3).

3.2.1 An Overview of the GDPR. The GDPR addresses a wide area of aspects that are related to data processing and privacy. In particular, it provides a lawful framework, specifying sanctions in case of infringement, and covers all steps that are concerned with the collection, usage, protection, and interaction with personal data in the EU. As such, the GDPR introduces notable changes to the privacy legislation in the EU.

EU-wide Law. Since directives are enacted through national laws, they contributed to considerably varying privacy standards throughout Europe [38]. The enactment of the GDPR aimed to remedy this situation by establishing high EU-wide privacy standards that regulate all processing of personal data of EU citizens, even if this processing takes place outside the EU [31, 147]. The previously established *right to be forgotten* [51, Article 17] and the aforementioned requirement to obtain explicit consent before personal data may be processed [51, Article 6(a)] have been put into EU-wide law as part of the GDPR, to name just two examples. Recital 10 of the GDPR specifically mentions that "[...] the level of protection [...] should be equivalent in all Member States", but also

leaves it up to the member states "[...] to further specify the application of the rules of this Regulation" [51]. Thus, member states still have the freedom to apply the GDPR more restrictively [143], i.e., the GDPR defines a minimum level of data protection and privacy in the European Union.

Binding Sanctions. In contrast to all preceding directives, the GDPR introduces binding sanctions that are imposed on service providers who fail to comply with its rules. Most notably, parties failing to comply with the requirements set by the GDPR face fines up to 4 % of their annual revenue or 20 million € [51, Article 83(5,6)]. In which specific cases such fines are appropriate, and how their exact amount is determined is also subject of the GDPR [51, Article 83(2,3,4)]. However, each EU state is granted the freedom to specify additional fines for infringements that are not specifically covered by the GDPR as long as the sanctions are "[...] effective, proportionate and dissuasive" [51, Article 84(1)].

Legitimate Interest. The GDPR includes multiple legal bases on which data controllers may carry out the processing of personal data. For this work, the statement that the processing is lawful if the data subject consented to it is especially relevant [51, Article 6(1)(a)]. Processing is also legitimate if it is required to fulfill contractual or other legal obligations, if it serves the public interest, or to protect vital interests or natural persons [51, Article 6(1)(b-e)]. Additionally, the GDPR introduces legitimate interest [51, Article 6(1)(f)] as a legal basis to process personal data. For example, the processing is legitimate if it serves the purpose of preventing or investigating fraud or other criminal activities [46]. To establish legitimate interest, the data controller must prove that their interest in processing the data outweighs the individual's interest for privacy.

Focusing on advertisements, the GDPR states that "[...] processing of personal data for direct marketing purposes may be regarded as carried out for a legitimate interest" [51, Recital (47)]. However, Recital 70 of the GDPR [51] states that every data subject must be provided the opportunity to reject such processing and that the processing party must make this clear to the data subject. The solution to these seemingly contradictory statements is that a party may rely on legitimate interest as a legal basis to process personal data (e.g., in the form of marketing), but is required to prove, among other factors, that the data subject can reasonably expect such processing to take place and the processing party's interest and risk to the individual's privacy are balanced [44].

A reoccurring example in literature where the processing of personal information, such as a physical address, is implicit and thus no explicit consent has to be provided, is that of a pizza shop [44, 155]. If a customer orders a pizza to their home address, they have to assume that the pizza merchant at least temporarily stores it to carry out the delivery. If the pizza shop then wishes to use this data further in the context of marketing, this constitutes a separate case where legitimate interest has to be established. Based on our research, the applicability of legitimate interest is not yet well understood [69, 78, 156], and may have to be further specified, e.g., through court rulings.

3.2.2 Legislation of the GDPR Impacting the Web. The GDPR addresses a wide area of aspects that are related to data processing and privacy in a general sense. As such, the GDPR does not limit itself to electronic data or even the web in particular. However, due to the ubiquity of the Internet, it inevitably affected online service providers. With the average European spending up to two and a half hours online every day [134] and facing about 17 trackers per website on average [42], the amount of personal data processed on the web is enormous. Beyond being able to reconstruct significant parts of individuals' browsing history by use of such data [83], even more sensitive information, such as religion, political attitude, and sexual orientation, may be inferred by processing it [20, 108] using, e.g., complex statistical analysis. Thus, the following provisions of the GDPR significantly impacted data processing on the web.

Right to Be Forgotten. Based on an EU court ruling from 2014 [82], the right to be forgotten was included in the GDPR [51, Article 17], requiring data controllers to delete personal data upon

the data subject's request. This obligation holds as long as at least one of the circumstances listed in Article 17 applies, most notably if the collection was based on the subject's consent or if it happened illegitimately. However, data controllers have the right to retain personal data if it was obtained on different legal bases. These bases include contractual obligations, e.g., in a business-to-business setting, or instances where personal data is processed in a transformative manner, e.g., anonymization [69]. Furthermore, if personal data is used, for instance, to train a recommender system, it may be technically infeasible to prove afterward whether this data concerned a specific individual [108]; thus, the system itself would remain unaffected by deletion requests [69].

Transparency. Apart from deletion, users of online services can claim their *right of access* to request their personal data from the respective service providers [51, Article 15]. Thus, the data subject has the right to know what information a data controller stores on them. To effectively make use of this right, the data subject has to be aware that such processing occurs. To this end, Article 13 of the GDPR requires data controllers to inform data subjects what data is processed, how long, by whom, for what purpose, and on which legal basis. Furthermore, this information must be provided clearly and comprehensively [51, Article 12(1)]. Therefore, excessively technical or bureaucratic formulations do not comply with the GDPR. Transparency further extends to data breaches, which leak personal information [51, Article 33]. If a data controller is implicated in such a breach, they have to disclose it to the responsible legal authority within 72 h, while the concerning individual only has to be informed if the leak poses a high risk to their personal "rights and freedoms" [51, Article 34].

Data Portability. If the processing of personal information by a data controller is based on the consent of the data subject, the data subject also has the right to obtain this information from the controller in a "common and machine-readable format" [51, Article 20]. This freedom is granted by the so-called "right to data portability", and it enables data transfer both between a data controller and data subject, as well as between two or more data controllers [29, 69]. The requirement to provide the data in a common, machine-readable format thus allows for it to be processed automatically and managed via digital interfaces, possibly through a third-party system [29].

Data Protection Officer. Furthermore, the GDPR warrants the appointment of a data protection officer [51, Articles 36-38] if a business's main activity involves handling personal data. A data protection officer has two main functions. First, this officer acts as the central point of contact for both individuals who have concerns regarding their personal data, as well as the supervisory authority of the respectively responsible government. Second, the data protection officer is tasked with supervising the privacy policies of their associated business and must ensure that the data controller is aware of potential breaches of legal provisions related to privacy.

3.2.3 ePrivacy Regulation. As of 2017, a bill, referred to as *ePrivacy Regulation*, mostly meant to update the ePrivacy Directive, is being discussed in the European Council [45]. While the GDPR introduces legislation on personal data, the ePrivacy Regulation should target all types of electronic communication to establish privacy and confidentiality. Originally intended to come into effect alongside the GDPR [22], it features more precise statutes regarding specific services and tracking technology [47]. Because the GDPR is not designed to specifically address electronic communication, most restrictions and obligations related to tracking on the web, in particular, are currently still specified by the ePrivacy Directive [135, 156]. This new regulation will most likely also require given consent for any interference on electronic communication [47]. However, due to a disagreement between EU member states, an enactment of the ePrivacy Regulation before 2025 is unlikely [22]. Even though this new regulation may be as disruptive as the GDPR, we consider it to be out of scope for this work, mostly because no aspect is fixed yet with reasonable certainty.

Takeaway. *The GDPR establishes precise requirements for data controllers, as well as penalties if a controller fails to comply with them. These requirements notably include, among others, that data subjects are informed clearly and comprehensively about the data processing they are subjected to, are provided the possibility to obtain copies of the processed data, and are given the right to retract their consent to the processing.*

3.3 Tracking Techniques and Privacy Policies

Nowadays, most web services employ some form of tracking [42] with online advertisement, a multi-billion dollar industry [56], acting as a major driver. The specific techniques of how personal data is collected and for what reason has significantly changed over the years [124, 129]. Similarly, the aforementioned changes concerning the legal framework regarding thus collected data have had a noticeable impact on web service providers and, consequently, the online browsing experience and the end-users' privacy. Therefore, in the following, we introduce different types of cookies (Section 3.3.1), cookie banners (Section 3.3.2), privacy policies (Section 3.3.3), and fingerprinting on the web (Section 3.3.4).

3.3.1 Different Applications of Cookies. Cookies are key-value pairs that may be used to persist configuration or identity data and constitute the most prominent form of online profiling, i.e., a way of identifying a specific end-user [102]. In this context, the question of whether such a profile can be linked to a natural person is irrelevant [69], as long as it represents a way of uniquely identifying an individual. Notably, this situation also demonstrates why simply pseudonymizing user data does not exempt it from regulation under the GDPR, which is an argument some providers bring up to legitimize their data collection practices [147]. In this subsection, we take a closer look at these practices and their uses. We are going to differentiate between two different types of cookies: *functional* and *non-functional* cookies. This distinction is based on the numerous different use cases and the fact that not all cookies constitute personal data. Our defined classes of cookies are loosely derived from related work (e.g., [31, 119, 149]) to give an intuitive notion of what they are used for in the context of this paper, rather than to provide a precise technical classification.

Functional Cookies. As HTTP, the protocol to commonly serve web services, is a stateless protocol, all information required to render a given website must be transmitted on the respective request [58]. This constraint often leads to the undesired effect of not being able to persist configurations across multiple requests natively, and constitutes the basis for most cookies. For example, an international website may set a preference cookie with a key *language* and values *English* or *German* to persist the information in which language a site should be served. In this case, the cookie is not considered personal data, since it cannot be used to identify a user. However, other examples of functional cookies may unambiguously identify individuals, for instance, if they are used to authenticate a user. For simplicity, we group all these cookies under the term *trackers*. To this end, after the user has logged into a web service, the web server sends a long random string to the browser, which is then returned with every subsequent request to prove that the same user makes these requests. To identify popular content of their provided service and be able to make specific recommendations to their users, service providers may want to collect data on their users' behavior and analyze it. Oftentimes, such an analysis is also enabled through cookies [149] that are uniquely identifying but of arguably functional nature [86]. An important note concerning functional tracking cookies is that the data subject must be informed that these cookies are used [51, Article 13], while explicit consent is not required, provided the data controller can prove that they are strictly necessary [86].

Non-Functional Cookies. Cookies that do not serve any functional purpose are more controversial. Non-functional trackers are often set by third parties to obtain browsing profiles of users and

thereby clearly identify a user on a given website. These kinds of cookies may be refused by a user without impairing the usability of the service in any way [51, Article 21]. Personalized advertisement is by far the most common use case for third-party trackers [21, 129, 145]. Additionally, third-party tracking naturally poses a much higher privacy risk to people browsing the web as it potentially comprises whole browsing sessions [36]. Furthermore, they leak information to third parties by nature and not just the accessed web service. The technology used to merge information obtained by multiple trackers from different sites is complex and ever-evolving [42, 150, 154], and most third parties share data among each other, enabling even more precise tracking on the web [42, 150]. In fact, this technology can also facilitate the generation of comprehensive browsing profiles without the need for third-party trackers [18, 19]. These practices can roughly be grouped under the term *cookie syncing* [42, 77, 150] and constitute a serious privacy threat for web users [42, 129]. One way to implement cookie syncing is by simply prompting the browser of the end-user to make a request to one tracking service with the cookie set by a different tracking service added as an additional parameter [150]. This way, the two parties can merge their tracking profiles. Unfortunately, the data on this form of tracking is much scarcer [42, 77].

3.3.2 Prevalence of Cookie Banners. Since the amendment to the ePrivacy Directive in 2009, websites commonly feature banners or popups to which we refer to as *cookie banners*. These banners inform users about the website's cookie policy and sometimes provide the option to reject certain cookies [149]. This noticeable change in the web landscape contributed to the situation that the 2009 amendment is sometimes called "Cookie Directive" [25] or "Cookie Law" [86]. In Figure 2, we illustrate the three most common types of such cookie banners [31, 74, 119, 149, 150]. Next, we introduce the three different types with increasing choice for the user in more detail.

Type 1. So-called cookie walls [24] (Type 1) only inform users that cookies are used and enable them to click **accept** to express consent. However, we can expect that cookies are being set regardless, whether the user clicks on the accept button or not [96, 119]. Since the user is not presented with an equal choice, websites with cookie walls cannot perform their data processing on the basis of explicit consent.

Type 2. The binary banner (Type 2) offers users the option to broadly reject cookies. The specific details of which kinds of cookies are prevented this way and which are still being set may, for example, be formulated in the privacy policy. According to Article 7 of the GDPR [51], consent can only be provided in a comprehensible manner so that the data subject knows what they consented to. Based on the research we inspected, after selecting **reject**, usually some cookies are still being stored [31, 74, 119] and **accept** allows all possible cookies [74]. Thus, it needs to be obvious for the user which cookies are necessary and cannot be rejected and which constitute trackers that are only set upon consent.

Type 3. We refer to multiple-choice (MC) banners as Type 3. They provide the user with more granular control of which kind of cookies they can reject. MC banners can provide different levels of granularity, ranging from a couple of broad categories (cf. Figure 2) to listing every tracking party individually [31]. However, tracking cookies are frequently set before the consent notice is displayed [96, 119], and instead of providing an opt-out mechanism of the personal data collection, the personalized content may be simply omitted [118]. Consequentially, the user either consents and violates their privacy or cannot fully use the service. The categorization of cookies in multiple-choice banners also varies and can be misleading, for example, by labeling cookies related to analytics used to personalize ads as necessary [31]. Thus, users cannot be certain that they are not being tracked even when selecting the most conservative cookie preferences.

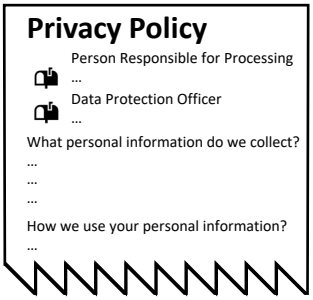
Additionally, cookie banners might be realized in a blocking manner, i.e., the user has to interact with it before the website can be fully accessed [128]. The level of obstruction can vary from site to

Fig. 2. Illustration of the three most common types of cookie banners ordered by an increasing amount of control provided to the user.

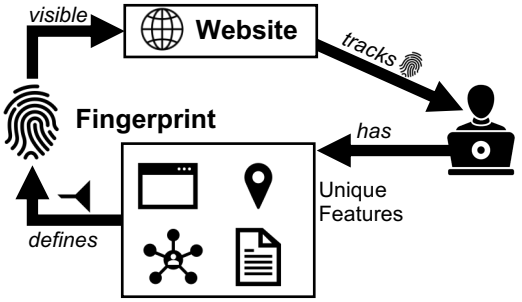
site. Such banners are more difficult for users to overlook and ignore [149], and thus urge users to either explicitly consent to or reject tracking. Another common observation is that broadly accepting all cookies is less complicated than broadly rejecting cookies through such consent notices [128, 149], which contradicts the GDPR's notion of equal choice [105, 128].

3.3.3 Evolution of Privacy Policies. As web services are mandated to disclose what kind of data is collected and to specify why this collection is performed in a clear and comprehensible manner, they usually offer *privacy policies* detailing this information. Large companies that host a multitude of web services, e.g., Google, commonly share the same privacy policy, and if third-party trackers are used, sites may link to the privacy policies of the respective service [31, 94]. Such policies are used to provide users with an overview of how services handle personal data (cf. Figure 3a). The volume, readability, and accessibility of privacy policies vary from service to service, but Jensen and Potts [80] found in 2004 that only one in about 400 browsing sessions even includes a visit to a privacy policy webpage, and Stabauer [132] found less than one in 600 users reading privacy policies in 2018. Furthermore, McDonald and Cranor [99] calculated in 2008 that the average user would spend approximately 200 h/yr. reading privacy policies if they had to read each one they accepted. This situation is reminiscent of terms of service agreements, which theoretically also require users to invest hours of reading before being able to use a service [92].

Different approaches for machine-readable privacy policies have been proposed over the years to mitigate this situation. *Platform for Privacy Preferences Project* (P3P) is the most notable and was developed by the *World Wide Web Consortium* (W3C) in the late 1990s and officially endorsed in 2002. However, it failed to gain wide adoption by service providers and web browsers [99]. The most central points of criticism are the voluntary nature and high complexity of the system [41]. A number of other proposed approaches, such as the *Do Not Track* header [39] and the *Enterprise Privacy Authorization Language* [9], exist. Today, none of them is widely used, mainly due to low adoption rates by web service providers [160]. Instead, several different browser extensions, such as *Privee* [160] and *PrivacyGuide* [138, 139], have been proposed and implemented to help users by providing short summaries of privacy policies. These extensions face two great challenges depending on their implementation. Either they rely on a user-maintained database of web services and their respective privacy policies, which demands volunteers to invest manual labor and thus do not scale well [139], or the information extraction is automated, thus having to deal with natural language processing which can result in error-prone summaries [160].



(a) An illustration of a typical, GDPR-compliant privacy policy. Information regarding the type of personal data that is processed, for which reason the processing takes place, and contact information of the data protection officer are mandated by the GDPR.



(b) A representation of how device fingerprinting enables the tracking of users on the web. The specific fingerprint may vary across different websites, since it can include different features unique to a particular device, which are requested by the website.

Fig. 3. Perhaps less apparent to the average user than measures, such as cookie banners, privacy policies and device fingerprinting are significantly linked to the GDPR as well. While privacy policies tend to increase the transparency for users of online services, fingerprinting provides a mean of tracking them without their knowledge. Thus, from a privacy perspective, understanding how such techniques work is beneficial for users.

3.3.4 Fingerprinting Techniques. Apart from all previously mentioned tracking techniques, the information about the configuration of the used web browser also contains enough information to uniquely identify users [37]. Boda et al. [12] showed that the "[...] font set, the time zones, and the screen resolution [...]" combined with a part of the IP address consistently identify browsing devices. This technique of tracking users without their knowledge is called *fingerprinting*, and due to its non-transparency, protecting against it is very difficult [90]. Laperdrix et al. [90] performed a comprehensive study of browser fingerprinting technology in 2020 and showed that even assessing the prevalence of fingerprinting is very challenging, with studies obtaining results from below 1 % [2] up to 68 % [3]. We provide a visual representation of this tracking technique in Figure 3b to give an intuitive understanding of how it works.

The browser-specific information which can be exploited to enable fingerprinting is partially required to guarantee that a service functions correctly, such as the user-agent header or the resolution of the device screen [90]. However, fingerprinting generally also requires an active collection of browser information, which is most commonly achieved by embedding JavaScript code into the website [3]. These scripts are frequently inserted by third parties [3], and the service provider itself may not even be aware of this practice [42]. Due to the nature of data that is used to enable fingerprinting, users may experience limitations regarding their browsing experience, if they wish to protect against it, e.g., by deactivating JavaScript [15] or fixing the browser window to a constant size instead of using full-screen mode. The Tor Browser, which enables higher privacy standards than other commonly used web browsers [159], used to not run in full screen by default and even specifically warned its users about leaking their screen resolution by activating full-screen mode. Firefox, one of the three most popular web browsers worldwide [133], can be configured to not send certain identifying settings as well [159], which can cause compatibility issues with websites [136]. Depending on the method used to generate fingerprints, determining whether fingerprinting is performed and how to protect against it accordingly is more or less challenging.

Further methods of tracking, for example, application cookies [124], exist and are actively used by web service providers [36]. However, these approaches are beyond the scope of this paper.

Takeaway. *Cookies and browser fingerprinting are two of the most common tracking techniques employed by advertisement services and constitute a severe privacy risk. However, discovering whether such tracking occurs on a given website is not always easy. If it does, different types of cookie banners may provide users with some control. Exhaustive descriptions of data processing carried out by a web service are usually provided through privacy policies, but rarely read by users.*

4 POST GDPR LANDSCAPE

Based on the aforementioned GDPR-induced requirements and restrictions related to data processing (cf. Section 3.2) and tracking techniques (cf. Section 3.3), we analyzed multiple studies to quantify the impact the GDPR has had on the handling of personal data within the World Wide Web [31, 60, 93, 94, 119, 148, 149]. To this end, we first highlight the challenges of collecting quantitative information (Section 4.1). We then group these studies by the quantitative measures they obtained and evaluate what kinds of changes they outline (Section 4.2), before focusing on the implementational challenges related to web services stemming from these changes (Section 4.3). Finally, we compare the obtained data to the legal requirements of the GDPR in an effort to ascertain how much it affected individuals' privacy, as well as web service providers (Section 4.4).

4.1 Challenges in Data Collection

To effectively measure the impact the GDPR has had on web services, we require objective metrics that allow us to rate any observed changes. Here, the heterogeneity of modern websites constitutes a major challenge that large-scale studies have to overcome. Two of the most commonly evaluated features of websites by research are privacy policies and cookie banners (cf. Table 1). Unfortunately, these aspects both present their own challenges in terms of data collection.

Privacy Policies. Privacy policies are especially problematic regarding automated analysis due to their high average complexity and ambiguous formulations [94, 139, 157]. Such formulations may be intentionally ambiguous, both in the case of policies of website providers as well as in legal documents, which lead to different interpretations even among scholars in the field [29]. Despite advances in natural language processing (NLP), the quality of such algorithms is very dependent on a large training corpus to effectively identify all policies when parsing webpages. For this reason, most researchers limit their analysis to English webpages [31, 94, 139, 157]. This restriction should be noted as it could potentially introduce a bias to the collected data, and thus its evaluation as the European Union consists of a multitude of different languages and jurisdictions.

However, before any analysis can be conducted, the first challenge is to implement a robust algorithm to automatically retrieve privacy policies from a website. Exclusively relying on anchor texts that contain the phrases *privacy* or *policy* is not always successful in practice [31, 160]. Web service providers may deliberately introduce other issues, for example, if infinite scrolling, i.e., dynamically adding new content below the visible area, is used to make links basically inaccessible [31]. These obstacles introduce a need for manual validation of the collected data [31, 42].

Trackers and Cookie Banners. Similarly, an automated approach of assessing the amount of control provided by cookie banners is almost impossible in practice, due to their differing designs [31, 119]. Thus, the analysis of cookie banners usually introduces a significant amount of manual labeling effort [31, 150], limiting the scope [149]. Although we found no evidence that measuring the number of trackers causes any considerable challenges, determining whether cookies are used for tracking is far from trivial [119]. Since cookies only consist of a key and a value that do not have to indicate their function, determining what they are used for is much more complicated [119, 129]. As a result, research may assume the worst case, thus gathering insight into the extent of tracking that is possible, rather than obtaining solid evidence of the prevalence of tracking. Especially, first-party tracking frequently serves a primarily functional purpose (cf.

Section 3.3.1), which is basically indistinguishable from tracking for other purposes [86, 90, 119]. Although tracing outgoing requests can lead to identifications whether a first party forwards user information to third parties, which also enables an analysis of cookie syncing [42, 77, 150], not all works we studied include such an analysis [25, 31, 74, 119].

Fingerprinting Techniques. Compared to the aforementioned privacy policies and cookie banners, which are usually even noticeable to uneducated users, fingerprinting generally happens unnoticed by users [90]. In a recent survey, Laperdrix et al. [90] analyzed a large corpus of scientific research regarding fingerprinting published over the last decade. They discovered that determining the prevalence of fingerprinting heavily depends on the interpretation of obtained information. For example, requesting the user agent and a list of installed fonts can be signs that a website is performing fingerprinting, but are also common practices to optimize a website for a specific device [90]. They also highlight that a lot of research regarding online fingerprinting is concerned with device-specific data that could be used to identify individuals in theory, but it is unclear to what extent such techniques are used in practice. A common approach to remedy this circumstance is to compare the obtained results and involved parties to a list of known tracking providers [31, 42, 145, 150]. Depending on the frequency with which these lists are updated, they may be incomplete and thus provide a lower bound for tracking analysis. For these reasons, browser fingerprinting constitutes one of the more difficult tracking techniques to measure.

Takeaway. *Today's responsive web design and inconsistent wording of privacy policies complicate automated measurements. Nevertheless, privacy policies and cookie banners are among the most studied aspects that directly concern the GDPR. When compared to studies of fingerprinting techniques, they also provide more consistent results across all research.*

4.2 Empirical Data

Despite the previously mentioned challenges concerning the data collection, a number of studies have collected and analyzed data around the time period the GDPR came into effect, showing which aspects of websites have changed the most and which parties were the most affected. The research we investigated puts special emphasis on the prevalence of third-party trackers [25, 42, 74, 81, 93, 119, 129] (Section 4.2.1), privacy policies [31, 79, 94, 119, 157] (Section 4.2.2), and cookie banners [31, 74, 105, 119, 128] (Section 4.2.3). We augment these technical aspects with a dedicated analysis of work concerning user behavior [88, 103, 149, 151] (Section 4.2.4) and security [17, 34, 60, 114] (Section 4.2.5).

In the following, we present what we consider to be the most meaningful findings in the respective work and point out similarities as well as differences between them. We provide a tabular overview of the respective measurements and results in Appendix B. One important aspect which is consistent within the featured research is given by differing results depending on both the country from which the website originates and the category of included content [31, 119, 129, 150]. Furthermore, the featured scientific works agree that countries and categories that performed worse prior to the GDPR were more affected by it than those that already provided good privacy protection beforehand, thus aligning privacy standards across the EU [31].

4.2.1 Third-Party Trackers. Even after the GDPR came into effect, as many as 92 % of websites may set uniquely identifying cookies as soon as the page is loaded [119]. At this point, a user has not yet been provided with any kind of cookie banner, privacy policy, or given the opportunity to express consent. A total of 80 % of these cookies also remain on the user's device for at least one year [119]. When comparing European to US websites, Dabrowski et al. [25] noticed in 2018 that out of the Top 500 websites (according to Alexa [5]), twice as many sites set persistent cookies for US (ca. 90 %) visitors as for users from the EU (ca. 45 %).

Websites focused on news content are especially interesting in the post-GDPR landscape as their financial model depends on third-party advertisement in large parts [42, 128]. Libert et al. [93] compared the number of cookies on news pages before and after the GDPR without the user's consent to tracking. Most third-party trackers on these sites originate from Google and Facebook, with a relative presence of 97 % and 96 %, respectively. These numbers dropped to 75 % and 70 % after the GDPR's enactment. Libert et al. [93] observed an average 22 % decrease of cookies with significant differences between countries, ranging from a 6 % decrease for German news sites to 45 % for sites from the UK. These findings match results obtained by Hu and Sastry [74], who observed a reduction of third-party cookies of around 18 % for the Top 100 websites that allow users to configure their consent, and Sørensen et al. [131], who report a decrease of up to 25 %. Urban et al. [148] additionally found that the amount of data sharing between parties decreased by 40 % after the enactment of the GDPR.

However, both Hu and Sastry [74] and Sørensen and Kosta [129] concluded that it is arguable whether the GDPR is the cause of the observed decrease in tracking. Although they found a significant decline of third-party traffic on webpages in the EU right after the GDPR was enacted, their collected data showed high fluctuation throughout and an upward trend since August 2018. They explicitly point out that the observed effects could also be the result of a change in tracking and advertisement technology and Urban et al. [148] tend to agree with this assessment. Especially when considering the browsing behavior of real users, the users' exposure to third-party cookies remains nearly unchanged [74]. The trend of Sørensen and Kosta's data also showed that by the end of 2018, third-party presence is almost back at the level it was before the GDPR was enacted [129]. This observation matches the results obtained by Johnson and Shriver [81], who state that the "[...] GDPR effect erodes over time". Furthermore, they, as well as others [21, 66, 148], highlight that the GDPR impacted smaller advertisement companies considerably more than large brands, such as Google and Facebook, leading to a higher market concentration for these companies, which, in turn, may increase the privacy threat, rather than decrease it.

4.2.2 Privacy Policies. When taking a look at privacy policies, Degeling et al. [31] observed that 85 % of websites updated their privacy policy before the GDPR took effect in May 2018, with the general trend being an increase in volume while showing a slight decrease in complexity [31, 94]. In particular, they report an increase of about 40 % from 2145 words in March 2016 to 3603 words in May 2018, for the average privacy policy within their dataset containing the Top 500 websites of all 28 EU member states, according to Alexa [5]. This finding is in line with the work by Linden et al. [94]. McDonald and Cranor [99] found the average privacy policy to be roughly 2500 words in 2008, which indicates that the volume of privacy policies decreased between 2008 and 2016. However, their dataset consisted of "[...] the 75 most popular websites based on [...] AOL search data" which could also account for the differing results.

Readability and availability are two common issues with privacy policies that Jensen and Potts [80] already identified as such in 2004. Even today, after the GDPR, they still remain problematic [31, 157], although the overall average complexity of privacy policies has measurably declined since then [119]. Linden et al. [94] concluded based on a user study that privacy policies overall improved in terms of readability. We combine Linden et al.'s [94] and Degeling et al.'s [31] findings and illustrate the corresponding changes in Figure 4a.

Results obtained by Wilson et al. [157] in a study from 2018 show that a regular, not specifically educated person from the US is able to identify whether a service engages in a given data-related practice based on its privacy policy with about 90 % accuracy. However, they also report that privacy policies from US websites, in the majority of cases, do not provide information about data collection and especially data sharing with sufficient clarity. Contrarily, Linden et al. [94] noticed

that policies from EU websites more specifically address the aspects of privacy that are relevant for the GDPR. Especially the amount of information regarding how long data is stored, by whom, and for what purpose increased. The results obtained by Fawaz et al. [55] support the notion that these changes can be attributed to the GDPR's enactment. At the same time, Bufalieri et al. [17] identify information on how users can obtain access to their personal data as a major blind spot among popular websites, with one out of five privacy policies featuring no information in this regard. Linden et al.'s [94] results also show that changes regarding the policies' semantic complexity are very small to statistically insignificant, maintaining a relatively high level of complexity for privacy policies from EU websites [119, 146]. Thus, the observed improvements in readability are overall not significant enough to conclude that privacy policies are understandable by the general public [119] and Linden et al. [94] pointed out compliance issues under the GDPR with 58 % of privacy policies of popular websites within the EU.

4.2.3 Cookie Banners. Degeling et al. [31] also investigated cookie banners and report an average increase from about 50 % to roughly 70 % on a dataset featuring over 9000 websites. This change is a significantly steeper increase of cookie consent notices than the 29 % increase measured by Miyazaki in 2008 [102], which can, for the most part, probably be traced back to the enactment of the ePrivacy Directive in 2002. Although EU sites offer a higher level of transparency regarding cookies [119] based on the prevalence of cookie consent notices, US sites were also affected by the GDPR [31]. Taking a look at the individual countries, the differences between the EU states are significant: with cookie banners on Slovakian websites only increasing by 14.6 % to as much as 70 % for websites with Irish top-level domains. Degeling et al. [31] also found that the vast majority of cookie banners (up to 90 %) are only informational (Type 1). However, this kind of cookie banner is often not conforming with the GDPR even if not interacting with it causes no cookies to be set, since the user is not presented with an equal choice to reject tracking [119]. Bornschein et al. [16] found 45 % of websites provide their users the opportunity to opt out of cookies, whereas Sanchez-Rola et al. [119] specify the number of websites with Type 2 and 3 cookie banners to be around 16 % in the EU and 12 % in the US. In this case, the difference may be due to the used dataset as Bornschein et al. [16] rely on a more limited and specific collection of websites, all within the scope of online commerce, and conduct their analysis three months later than Sanchez-Rola et al. [119]. Sanchez-Rola et al. [119] specifically investigated the effectiveness of cookie banners and found that they completely stop tracking in only 3 % of cases, which might also account for that discrepancy. They discovered a clear correlation between a reduction in tracking and the type of content the website hosts. On the one hand, websites serving pornographic content perform among the worst when it comes to respecting users' rejection of trackers, next to sites advertising or selling weapons, and religious content. Government websites, on the other hand, perform the best. The popularity of a website also factors into the amount of control. Hu and Sastry [74] identified the percentage of Type 3 cookie banners among the Top 20 most visited sites to be as high as 70 %. They further point out that a correlation between the amount of control websites offer and the maximum amount of cookies set might exist, with more control also implying more potential trackers. Thus, the most popular websites that are especially attractive to advertisers and feature more third-party content provide their users with a higher than average amount of control which cookies they allow. A study by Hils et al. [73] confirms these results, finding that adoption of multiple-choice cookie banners is most prevalent in the Top 10 000 of most popular websites. One possible explanation for this is that through the increased control, service providers hope to also increase their users' trust in their privacy protection measures [16, 80, 102].

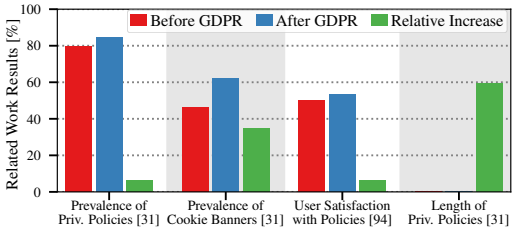
Since the introduction of the GDPR, the prevalence of cookie banners provided by so-called *Consent Management Platforms* (CMPs) has sharply increased [73]. CMPs are essentially code

libraries that can be integrated into an existing website to process and manage user data which may be done for a variety of purposes [73]. As the name suggests, an important feature of CMPs is their management of user consent which is achieved through various types of cookie banners [73, 96, 105]. Among the Top 10 000 most popular websites, the adoption of CMP banners has risen from less than 1 % at in 2018 to almost 10 % at the end of 2020 [73]. The general consensus on studies investigating cookie banners in detail is that, whether provided by CMPs or not, they are not fully compliant with the requirements set by the GDPR in the majority of cases [96, 100, 105, 128, 149]. The two most pervasive problems are pre-ticked checkboxes [96, 105] and nudging users to accept all cookies [128] by either making the opt-out settings less accessible than opting in or requiring an additional confirmation for opting out but not to opt-in. Pre-ticked checkboxes are especially contentious, since this practice was specifically outlawed by a recent court ruling on the basis of the ePrivacy Directive [23, 156].

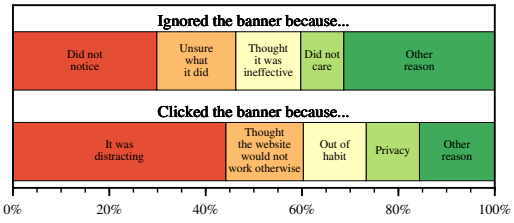
4.2.4 User Behavior. A user study conducted by Utz et al. [149] had users randomly interact with different types of cookie banners and captured the responses in a survey. We detail the reasons users gave for interacting or ignoring the cookie banner in Figure 4b. The general conclusion we can draw from this survey is that users overall do not agree with cookies, but are often unsatisfied with consent notices or insecure about the effect they have. As a result, they commonly just ignore cookie banners (around 70 % of the time) if possible [149].

For example, on the one hand, 29.7 % of participants thought that the content of the website could not be accessed, unless they accepted cookies for this website. On the other hand, 12.8 % answered that they do not think the cookie banner prevents data from being collected. The same percentage also stated that they did not think any cookies would be stored on their browser if they decline, but only 11.7 % answered that they believe no personal data is collected. Most participants stated that their interactions with cookie banners are "privacy-focused", meaning that they always prefer the "least invasive option" [149]. Based on another study from 2019, Urban et al. [146] reported that online users have an incomplete understanding of both tracking and their own privacy rights, such as the right of access. A comparable study performed by Kulyk et al. [88], before the GDPR's enactment, describes similar user behavior: Users are mostly annoyed by cookie banners, but at the same time, they feel like their privacy is threatened by cookies, especially if they are informed that the collected data is used for advertisement. Varkonyi et al. [151] found in a survey six months after enactment of the GDPR that less than half (47.3 %) of a multinational group of 110 university students, mostly enrolled in either law or computer science, knew what the GDPR was. In a different survey carried out by Mohallick et al. [103] three months before the GDPR went into force, 200 researchers were asked about their attitude towards privacy and the role of data controllers. The main concern based on the results was that data controllers might share personal data with third parties (66 %), which is, in fact, a common practice (cf. Section 4.2.1). However, participants also stated that they would be more trustful if data controllers asked permission before processing (77 %) or provided the opportunity to edit and delete data (66 %).

By comparing the prevalence of ad-blockers given by Pujol et al. [110] from 2015 and Utz et al. [149] from 2019, we can make a rough estimate regarding the adoption rate of software blocking trackers around the time the GDPR came into effect. In 2015, the ad-blocker usage amounted to an average of 19.7 % among Europeans [110], which is only marginally less than the 20 % measured by Utz et al. [149] for Europeans in 2019. Pujol et al. [110] point out the difficulties in obtaining reliable data concerning ad-blocker usage, and thus the lack of a significant change in tracker blocking software may be due to measuring errors or limited datasets. However, Utz et al. [149] state that the prevalence of ad-blockers they observed is consistent with numbers from 2017, and thus, ad-blocker usage was likely not significantly impacted by the GDPR in either direction.



(a) Relative changes in privacy policies and cookie banners after the GDPR was enacted. Related work reports an increased prevalence of privacy policies and cookie banners following the GDPR.



(b) The reasons for ignoring or clicking cookie banners are diverse, as data taken from Utz et al. [149] suggests. Around 70 % of the time, cookie banners are simply ignored by the users [149].

Fig. 4. Composition of different metrics affected by the GDPR. Privacy policies have become significantly longer, and cookie banners more common, but user studies show that many of them are critical of the changes and are generally not satisfied with the current situation.

4.2.5 Security. Our corpus also includes a small number of studies related to security. A small scale-study of 33 online academic repositories by Formanek et al. [60] found only half the amount of sites using HTTP instead of HTTPS in 2018 compared to 2016 and that the cryptographic strength of the used encryption almost doubled. The rest found either no significant change in security [114] or additional security challenges due to the GDPR's right of access [17, 34]. The main issues in this regard are that personal data is transmitted in an insecure manner in 30 % of cases [17], and Di Martino et al. [34] were able to obtain full access to personal data records by impersonating the data subject in 15 out of 55 cases (27.27 %). The insecure ways of transmitting personal information observed by Bufalieri et al. [17] were either data being sent as unencrypted email attachments or the decryption key getting sent to the same email address.

To summarize our analyzed literature, in Table 1, we provide an exhaustive list of studies collecting measurements relating to the GDPR that we analyzed in the course of this work. In some instances where no clear date was given when the collection took place, we instead list the earliest date of publication of the respective work we could find.

Takeaway. *The reported data shows that since the GDPR has been enacted, the amount of third-party tracking has declined, and transparency and control on the side of the users have improved overall. The prevalence of cookie banners and privacy policies and their volume has significantly increased across all datasets leading to more consistent privacy levels on EU websites. However, the relatively low ratio of multiple-choice cookie banners and relative high complexity of privacy policies remains an issue as users, in general, prefer low-effort privacy-preserving choices. Additionally, pinpointing the reason for observable changes in third-party tracking is challenging as trends show that the amount of trackers is close to pre-GDPR levels after a temporary decline.*

4.3 Implementational Challenges

Ensuring complete compliance with the GDPR poses a considerable implementational challenge in many cases [31, 146], at least if we assume that services cannot limit their existing data collection practices. As such changes might negatively impact the user experience or the web services' revenue [16], the web service providers instead had to implement far-reaching changes to their services in light of the GDPR. Here, the main problem is that a lot of data collection happening on websites, potentially violating user privacy, is automatically conducted by third parties [78, 96, 107, 112, 119, 153]. The issue with third parties in this context comes from the fact that even if

Author (Year)	Type of Measurement	Period of Measurement
	Data Protection Directive (1995)	
	ePrivacy Directive (2002)	
Miyazaki (2008) [102]	Cookie usage & cookie banners	Jun. 2000 & Feb. 2007
Jensen and Potts (2004) [80]	Privacy policies	Jul. 2001 & Sep. 2003
McDonald and Cranor (2008) [99]	Length of privacy policies	2008
	Cookie Amendment (2009)	
Acar et al. (2013) [2]	Browser fingerprinting	Nov. 2013
Wilson et al. (2018) [157]	Privacy policies	Dec. 2013 & Jan. 2014
Rao et al. (2016) [113]	User study on privacy policies	Feb. 2015
FaizKhademi et al. (2015) [54]	Browser fingerprinting	Jun. 2015
Pujol et al. (2015) [110]	Ad traffic & ad blocker usage	Oct. 2015
	General Data Protection Regulation (2016)	
Englehardt and Narayanan (2016) [42]	Tracker usage	Jul. 2016
Cabanas et al. (2018) [20]	Ad-personalization on Facebook	Oct. 2016 – Oct. 2017
Sørensen and Van den Bulck (2020) [130]	Third-party content	Dec. 2016 & Apr. – Aug. 2017
Rao and Pfeffer (2019) [112]	Tracker flow	Apr. – May 2017
Mohallick et al. (2018) [103]	User study on privacy	May 2017 – Feb. 2018

Table 1. Overview of related work with quantitative measurements. We sorted the work according to the measurement period and also include the type of data that the authors analyzed. For temporal context, we also include the EU privacy legislation. Please note that we do not group the work w.r.t. this legislation.

Table continues on the next page.

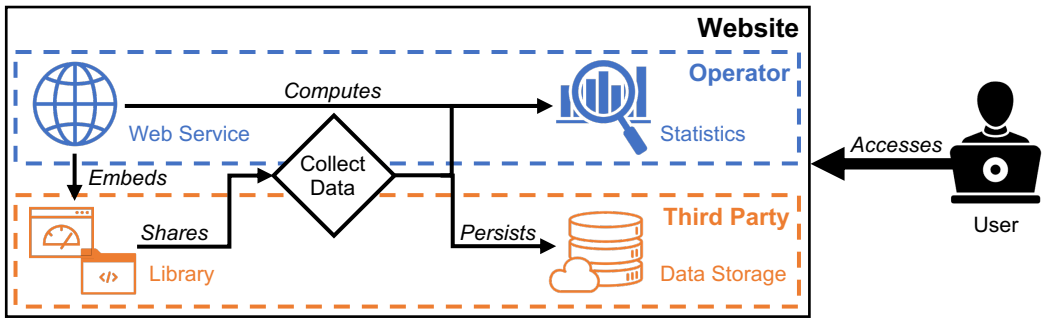


Fig. 5. The dataflow of a website using a third-party analytics library shows that personal data is also persisted by this external service provider that is not under the control of the website operator. Thus, the first party operating the web service has no direct access to it.

third-party trackers are only set after a user has given consent, the GDPR requires that consent can be withdrawn, and thus, the obtained data has to be deleted. Since the data is not stored by the party running the website, they have no direct access to it and can only report it to trigger a deletion. However, for this functionality, the third-party must provide an appropriate interface. In Figure 5, we illustrate how third-party code libraries can lead to a loss of control over personal data for both users and website operators. Such libraries provide diverse functionality, such as data collection (as shown in Figure 5), authentication, social media integration, and other features, but the primary purpose of third-party content remains advertising [129, 130].

The probability that a given site takes part in third-party tracking and how many third parties are present on a website heavily depends on the type of content hosted on the site [129]. Thus, different researchers have obtained various results regarding the prevalence of third-party tracking

Author (Year)	Type of Measurement	Period of Measurement
---------- **GDPR enforced (applicable)** (May 2018) ----------		
Trevisan et al. (2019) [145]	Cookie usage & cookie banners	Jan. 2015 – Nov. 2018
Dabrowski et al. (2019) [25]	Tracker usage	2016 & 2018
Linden et al. (2020) [94]	Privacy policies	Jan. 2016 – May 2019
Degeling et al. (2019) [31]	Privacy policies & cookie banners	Mar. 2016 - May 2018
Formanek et al. (2019) [60]	HTTPS adoption and security	Jun. 2016 & Jun. 2017 & Jun. 2018
Goldberg et al. (2019) [66]	Website revenues	Jan. – Sep. 2017 & Jan. – Sep. 2018
Stabauer (2019) [132]	User study on cookie banners	Jan. 2017 – Nov. 2018
Sakamoto and Matsunaga (2019) [118]	Tracker opt-out behavior	Jul. 2017 & Jun. 2018
Fawaz et al. (2019) [55]	Privacy policies	Jul. 2017 & Jan. 2019
Iordanou et al. (2018) [77]	Tracker usage and flow	Sep. 2017 – Jan. 2018
Raponi and Di Pietro (2020) [114]	Password security	Dec. 2017 & Dec. 2018
Vallina et al. (2019) [150]	Third-party content	2018
Hu and Sastry (2019) [74]	Tracker usage & cookie banners	Jan. 2018 – Jan. 2019
Sørensen and Kosta (2019) [129]	Third-party content	Feb. – Sep. 2018
Sørensen et al. (2020) [131]	Third-party content	Feb. 2018 & Oct. 2019
Hils et al. (2020) [73]	Cookie banners	Mar. 2018 – Sep. 2020
Cliqz-GmbH (2018) [21]	Tracker usage	Apr. – Jul. 2018
Libert et al. (2018) [93]	Third-party content	Apr. – Jul. 2018
Kulyk et al. (2018) [88]	User study on cookie banners	Apr. 2018
Johnson and Shriver (2019) [81]	Third-party content	May – Dec. 2018
Vlajic et al. (2018) [153]	Third-party content	May – Jun. 2018
Urban et al. (2020) [148]	Data sharing	May 2018 – Mar. 2019
Urban et al. (2019) [147]	Right of access responses	Jun. & Sep. 2018
Sanchez-Rola et al. (2019) [119]	Privacy policies & cookie banners	Jul. 2018
Al-Fannah et al. (2018) [3]	Browser fingerprinting	Aug. 2018
Latham and Goltz (2019) [91]	User study on privacy and AI	Oct. 2018
Boniface et al. (2019) [13]	Right of access authentication	Oct. 2018
Schmidt et al. (2020) [123]	User study on cookie banners	Oct. & Nov. 2018
Di Martino et al. (2019) [34]	Right of access security	Oct. 2018 – Mar. 2019
Bornschein et al. (2020) [16]	Cookie usage & cookie banners	Nov. 2018
Varkonyi et al. (2019) [151]	Survey on GDPR	Nov. 2018
Utz et al. (2019) [149]	User study on cookie banners	Nov. 2018 – Mar. 2019
Jakobi et al. (2020) [78]	Tracker usage	Feb. 2019
Urban et al. (2019) [146]	Right of access & user perception	Feb. 2019
Bufalieri et al. (2020) [17]	Right of access security	May 2019
Mohan et al. (2019) [104]	Privacy policy compliance	Jun. 2019
Nouwens et al. (2020) [105]	Cookie banners	Sep. 2019
Matte et al. (2020) [96]	Cookie banners	Sep. 2019
Borgolte and Feamster (2020) [15]	Tracker usage	Sep. & Oct. 2019
Soe et al. (2020) [128]	Cookie banners	Jul. 2019 & Apr. 2020
Weinshel et al. (2019) [154]	User study on tracking	Nov. 2019
Javed et al. (2020) [79]	Privacy policies	Mar. – Jul. 2020
Mehrnezhad (2020) [100]	Cookie banners	April 2020

Table 1. Overview of related work with quantitative measurements. We sorted the work according to the measurement period and also include the type of data that the authors analyzed. For temporal context, we also include the EU privacy legislation. Please note that we do not group the work w.r.t. this legislation. On the one hand, we can observe a heavy focus on third-party content, (tracking) cookie usage, and other forms of tracking around the enactment date of the GDPR. Privacy policies and cookies banners, on the other hand, can be studied over longer periods of time using tools such as the WaybackMachine [140]. However, relevant research datasets regarding cookies, privacy notices, and even fingerprinting techniques also date back multiple years prior to the GDPR.

ranging from around 70 % [81, 150] up to around 80 % [119]. The average amount of unique third parties for a given website is by far the highest for news websites, with approximately 30 parties per site on average [129], whereas the overall average is somewhere between 12 [81] and 20 [74, 129]. The datasets analyzed in these cases were all obtained by collecting the Top 500 [74, 129] up to Top 2000 [81] European websites, according to Alexa [5]. Thus, website hosts of popular online services are clearly struggling with becoming independent of third parties. Nevertheless, a positive example of how compliance with the GDPR can be achieved is given by the Google Analytics library, which includes the function to delete trackers set by it on Google's side [67]. While CMPs often provide the necessary means to opt out of third-party tracking [31, 96], their adoption rate is rather low, with about 3 % [119] to 6 % [96] of websites implementing them. Furthermore, CMPs seem to be regularly configured incorrectly, or only parts of the provided functionality is utilized [96]. Since the website developers create the classification of "strictly necessary" cookies, improperly categorized cookies remain an issue concerning GDPR-compliant cookie policies [31]. Although our research does not allow for a conclusive statement, the authors of various works conjecture that the reason may be a fear of negative effects on the revenue mixed with weak enforcement of the GDPR [16, 105, 119]. Multiple authors [16, 128, 149] point out fundamental discrepancies between the current business model of online services and the amount of control over personal data required by the GDPR. Utz et al. [149] conclude, based on a study, that GDPR-conform notices would result in user acceptance rates of less than 0.1 %. Thus, the biggest challenge may be more economical in nature than technical.

Besides limiting tracking, online services are also burdened with having to process additional user requests on the basis of the GDPR [146]. These requests include users asking for copies of their personal data stored by data controllers, demanding rectification or deletion of such data, and other inquiries to be handled by a data protection officer. Di Martino et al. [34] and Boniface et al. [13] demonstrated that many data controllers request additional information, such as names, phone numbers, home addresses, or copies of ID cards, before making stored data available, which leads to a leakage of personal data. On the one hand, this situation was most likely not intended when enacting the GDPR, but on the other hand, research suggests that more involvement by data protection authorities could improve this deficiency [13, 146]. Furthermore, although around a ⅓ to ½ of the large online service providers implement at least partially automated systems to respond to users' privacy concerns [34, 147], every data controller is still required to maintain a human point of contact in the form of a data protection officer. Since the respective requests (including the right to data portability) have to be answered within a timespan of 30 days, the workload of data controllers increased due to requirements introduced by the GDPR [33]. However, the amount of data subject requests fell short of the expectations of 42 % of data controllers interviewed in a study by Urban et al. [146].

Takeaway. *A major challenge in achieving GDPR-compliant data policies results from web services' dependence on third-party trackers, which is oftentimes economic in nature. However, technical hurdles, such as implementing efficient systems to respond to users requesting to acquire copies of or to delete personal data, introduce an overhead as well.*

4.4 Comparison against GDPR Requirements

Considering the observations, which we presented in Section 4.2, and the legal information that we discussed in Section 3.2, we can conclude that the majority of web services have not applied enough changes to their personal data policies to be fully compliant with the GDPR. Although most websites updated their privacy policy notices, thus increasing transparency, and the use of cookie banners has increased significantly, the necessary control over personal data is rarely provided to users. This situation may be due to disagreements between service providers and consumer

protection authorities concerning the interpretation of the GDPR, exemplified by recent rulings by European authorities.

The Dutch Data Protection Authority released a statement regarding cookie walls (Type 1, cf. Section 3.3.2) after receiving complaints by web users [10], which clearly states that based on the GDPR, this type of cookie consent notices does not protect the usage of trackers. Furthermore, the German Federal Court of Justice ruled that pre-ticked checkboxes do not comply with the ePrivacy Directive, which was prompted by a challenge by The German Federation of Consumer Organizations [23]. Based on the sources we analyzed [29, 38, 69, 135, 156], we believe that current and future court cases will clearly shape the boundaries for service providers set by the GDPR in the coming years.

As a matter of fact, the amount of legal actions against website providers based on violations of personal data protection laws has increased [120]. For example, Google is currently facing a fine of 50 million € due to a lack of transparency and prior consent regarding personalized advertisement [117]. A German online food delivery portal had to pay a fine of close to 200 000 € due to illegitimate storage of personal data and advertisement practices [26]. The UK's Information Commissioner's Office (ICO) filed a complaint against *The Washington Post* based on the personal data policy of their online newspaper outlet, which until this point in time has not yet provoked any change or reaction [127]. Here, users had the opportunity to either pay 90 $ for a year-long subscription or implicitly agree to tracking on The Washington Post's website. In a case involving physically collected data, an 18 million € fine was imposed on the Austrian mail service [8], due to the unlawful processing and sharing of private data. Other examples of legal cases resulting in fines either involve the publishing, sharing, processing, or retaining of personal data contrary to users' wishes [7, 27], only a few of them relate to online data traffic.

Another possible cause for legal actions stems from data breaches that are also subject to the GDPR [51, Article 33]. The number of reported instances increased significantly shortly before the GDPR took effect in 2018 [14], illustrating the raised demand for security as well as privacy and transparency. Data controllers need to provide a reasonable amount of security, for example, through the use of encryption, which is a debatable wording because what qualifies as reasonable is unclear [57]. However, if personal data gets stolen or leaked, the data holder has to make sure to properly disclose this fact [51, Articles 32,33]. The breaches involving British Airways [121] and the Marriott hotel group [116] constitute two high-profile cases where data controllers failed to comply with these new obligations. British Airways and Marriott were fined 230 million $ and 124 million $ respectively under the GDPR. Comparing these numbers to the circa 0.5 million $ fine against Facebook imposed by a British court due to the data breach involving Cambridge Analytica [142], which occurred before the GDPR's enactment, data controllers now face considerably higher fines if they fail to protect personal data.

Takeaway. *The GDPR provides the basis for many legal cases dealing with the processing of an individual's private data, triggered by a lack of transparency, control, or security. Furthermore, European authorities have released clarifying statements, forbidding cookie walls and pre-ticked checkboxes on consent prompts, further shaping the GDPR's impact.*

5 DISCUSSION AND OUTLOOK

The presented findings in Section 4 provide valuable insights into the (empirical) impact of the GDPR. A key aspect most work identifies is that although the majority of service providers have taken some actions to increase their users' privacy, the amount of personal data collected on websites still remains high. As a result, a number of high-profile legal cases surfaced following the introduction of the GDPR. However, these effects fall short of the dramatic effects anticipated by some experts [4, 72]. In this section, we mainly put the findings from Section 4 in a broader

context addressing claims from both scientific and journalistic resources. We further discuss the most intriguing points of contention and uncertainty related to the GDPR from our point of view.

Here, we deem the following aspects as important: the limitations of both effectively measuring the influence the GDPR has had on the web (Section 5.1), as well as the GDPR itself in terms of guaranteeing an individual's privacy (Section 5.2). In this context, we also take a closer look at the different parties affected by the GDPR and point out the challenges and opportunities it holds. Furthermore, we address data portability (Section 5.3), i.e., Article 20 of the GDPR, which constitutes an avidly discussed provision of the GDPR and its possible consequences for web services and their users. Finally, we elaborate on the notion of privacy by default (Section 5.4).

5.1 Missing Quantitative Results and Metrics

As we compiled in Section 4.2, the amount of accessible quantitative data regarding the impact of the GDPR on web services is relatively scarce. A major issue we observe is the diversity of measurement approaches which severely limits the comparability of pre-GDPR to post-GDPR research. While we can observe some trends, for example, we know that privacy policies have become longer and feature more of the keywords related to the GDPR [94], we are missing clear data on the number of services that fully comply with the GDPR. Obtaining reliable data relating to this issue is extremely time-consuming [157], and practically impossible to automate due to the variety of web services and amount of aspects to consider. Another obstacle is that even scholars have different interpretations of the GDPR's wording [29, 84]. Thus, both web service providers, as well as their users, face difficulties assessing whether a particular service is compliant with the GDPR or not. This circumstance is unsatisfactory for users, since they may feel that their privacy is not respected, but they are unsure whether legal action is warranted. Similarly, service providers may also feel a lot of pressure, since they are incentivized to collect as much data as they may to improve their revenue [20, 78]. Thus, they might interpret privacy-protecting legislation more loosely than originally intended, potentially risking severe fines. Cookie Banners function as a good example in this regard. Degeling et al. [31] argue that Type 2 and some Type 3 banners are not GDPR-compliant as they generally do not provide the user with sufficient information regarding the nature and purpose of data processing they might agree to. Since these banners lack clarity, which kinds of cookies are being set when pressing any of the buttons (as this information is rarely provided [31, 149]), this requirement is frequently not fulfilled.

Fingerprinting. Based on the work we analyzed, the amount of information regarding the specific methods employed to obtain personal data contained within privacy policies is also not well covered by current research. Fingerprinting is particularly interesting in this regard, as it specifically enables tracking without making it noticeable for the user [135]. Several studies have shown that browser fingerprinting is not an unusual practice [2, 3, 42, 54]. However, they disagree on the exact numbers. The impact of the GDPR on these techniques has, to the best of our knowledge, not been the subject of scientific research. We assume that the main reason why no such research exists is probably that obtaining data on the prevalence of fingerprinting from before the GDPR was passed after the fact is very difficult. Another issue regarding fingerprinting is the uncertainty of whether the GDPR prohibits first-party browser fingerprinting without user consent [135, 144].

Furthermore, comparing studies from different time periods investigating browser fingerprinting is problematic as the classification of fingerprinting varies significantly, resulting in majorly different conclusions [90]. For example, Engelhardt and Narayanan [42] found that around 5 % of the Top 1000 websites likely employ fingerprinting, while Al-Fannah et al. [3] claim in their study from two years later that "[...] 68.8 % of the Top 10 000 websites are potentially engaged in fingerprinting [...]". As they do not provide a specific date for when the data collection took place, we cannot be sure about their measurement period. However, Al-Fannah et al. [3] published their work in August 2018,

after the GDPR took effect. Mehrnezhad [100] obtained very similar results in 2020, while Vallina et al. [150] in 2018 are more in line with Engelhardt and Narayanan [42]. Thus, repeating studies similar to the mentioned studies [2, 3, 42] and comparing the findings would provide insights into how fingerprinting technology has evolved since the GDPR. Here, the qualification methods should be comparable to the methods of previous work. However, given that fingerprinting technology is constantly evolving [124, 135], the qualification may have to be adjusted to accurately reflect the prevalence of tracking through today's fingerprinting technology. These factors make browser fingerprinting much less comparable over a long period of time, i.e., giving a reasonable estimate of how much the GDPR affected this tracking technique is almost impossible.

User Interaction. Assessing the users' opinions and observable actions is another aspect that is very difficult to assess. On the one hand, some indicators that the average level of privacy awareness has increased considerably in the past are the number of legal actions against illegitimate practices [7, 27, 120] and the increased willingness to report privacy breaches [14]. However, we were unable to find a study that conclusively investigated users' impression of the changes brought about by the GDPR. The main problem with data relating to user behavior is that scientific user studies are, in most cases, limited in their sample size [94, 149]. In the studies we looked at, there was also a slight overrepresentation of young, European, and highly educated participants [91, 103, 146]. Results obtained by Bornschein et al. [16] indicate that this group is on average more privacy-aware. Thus, we have to be careful with generalized statements regarding online users. Service providers with access to large amounts of user data, on the other hand, usually do not remain neutral in their analyses. For example, one of the leading consent management platforms [73], Quantcast, reports a 90 % user acceptance rate [111]. However, most users report the desire to not be tracked [105, 149], but also tend to ignore cookie banners [149]. If we again consider the fact that websites regularly interpret no action as consent and only limit trackers once a user explicitly objects [119], many users might still be tracked without being aware of this practice. In fact, Hue and Sastry [74] postulate that frustration with cookie banners might cause them to simply consent to tracking, since it usually requires less effort, which in turn accounts for the observed steady rise in third-party tracking, following the initial drop-off after the GDPR's enactment.

Data Sharing. Besides frustrated users, we also came across a different hypothesis explaining the reduction in third-party tracking around the time the GDPR was enacted, shown by various studies [66, 74, 93, 129]. Multiple authors relate these changes to a potential shift in tracking and advertisement technology [129, 148], but point towards a need for more research regarding the data sharing practices employed by tracking services. Since "[...] the effects of the GDPR might not be directly measurable", Urban et al. [148] suggest that approaches to gain valuable insight in this regard may be to "[...] actively [make] use of the right to access [...] or by conducting expert interviews". Indeed, we estimate researchers' ability to measure data traffic between two data controllers as severely limited (if at all existent), since the studies we analyzed almost exclusively rely on client requests to web services. Goldberg et al. [66] obtained a dataset of actual tracking data from an analytics service provider. On the one hand, including such datasets in the analysis of online tracking might provide additional insights into the impact of the GDPR on the quality of tracking data. On the other hand, this does not necessarily provide additional details on how tracking data is obtained.

Takeaway. *Some changes in online tracking, such as the role of fingerprinting and exchanging of personal data, potentially brought about by the GDPR may not be easily measurable from the outside. Repeating pre-GDPR studies with careful consideration of the methodology, analyzing tracking data from analytics services, or conducting insider interviews may yield valuable insights into post-GDPR tracking on the web.*

5.2 Limitations of the GDPR

Due to the limited amount of objective data, the implications of the GDPR are a controversially discussed topic, including a lot of speculation. While some argue that the GDPR could or already has caused significant damage to a web user and service providers [66, 72, 109], others claim that through it, online users' privacy improves dramatically [4, 135], and yet others conclude that the GDPR allows for multiple interpretations regarding certain aspects [29, 38].

GDPR Coverage. Szymielewicz and Budington [135] argue that the ambiguity in the GDPR is intentional and necessary for it to be applicable to privacy breaches regardless of the involved specific technology. Then again, the lack of clarity also offers accused parties some leeway to argue that the GDPR does not specifically forbid their particular practices. Gray areas, such as browser fingerprinting and the applicability of legitimate interest, result from this situation that may be cleared up through future legislation [45]. In some instances, web service providers simply claimed legitimate interest in all of their tracking techniques [31, 119], rather than applying any changes. Unfortunately, we cannot provide a general statement whether such practices are lawful or not, because the boundaries of legitimate interest are especially fuzzy when dealing with web analytics [28]. Laperdrix et al. [90] conclude that browser fingerprinting clearly requires user consent, while Krausova et al. [87] argue that data obtained through browser fingerprinting cannot be considered personal data. Even assuming the most restrictive interpretation of the GDPR, proving that personal data is collected, stored, and processed remains difficult. Web browser fingerprints do not always allow for definitive identification [2], and proving whether advertisements are personalized may be unfeasible [108].

What is more, taking legal action is a costly and potentially risky endeavor, making it doubtful that individuals will bring every offense to court. Certainly, the press coverage resulting from such instances could improve the awareness of privacy overall. A study performed by Weinshel et al. [154] showed that the average user underestimates the extent to which they are being tracked on the web. Additionally, Al-Fannah et al. [3] claim that 88 % of users do not take any active precautions to mitigate susceptibility to browser fingerprinting. A reason might be that consistent protection from privacy threads through browser fingerprinting on the client-side alone is generally not possible without impeding the browsing experience [90]. Thus, the deterrent of significant fines and consumer awareness may provide better protection against browser fingerprinting.

Legal Boundaries. We have to consider here that the GDPR does not implement particularly strong control mechanisms itself apart from the data protection officer, who is supposed to act as an intermediary between interests of technical, economic, and privacy-related matters in a given company [38]. Unfortunately, the priority placed on the role of data protection officer in practice is usually rather low [33, 104]. Consequently, consumer organizations and communities, such as the Electronic Frontier Foundation [40] and NOYB [106], constitute important agents of individual privacy rights, as they are commonly the initiating party of legal trials [7, 27, 120].

Another challenge in this regard tackles the question of jurisdiction. While EU courts can sometimes leverage treaties to prosecute foreign parties [127] or may cease their EU-based assets [126], the case against The Washington Posts illustrates such actions are not always possible. After a reader from the UK alerted the ICO, the UK's main data protection authority [76], they filed an official complaint against the newspaper's tracking policy (cf. Figure 6) [127]. However, a spokesperson of the ICO stated that "[...] if [the Washington Post] choose[s] not to [heed our advice], there is nothing more we can do in relation to this matter" [127]. This statement reveals a significant limitation of the GDPR, due to the considerable amount of traffic from the EU to US-based web services. For example, Iordanou et al. [77] report that about 10 % of data traffic related to online advertising leaves the boundaries of EU jurisdiction, mostly with destinations in North America.

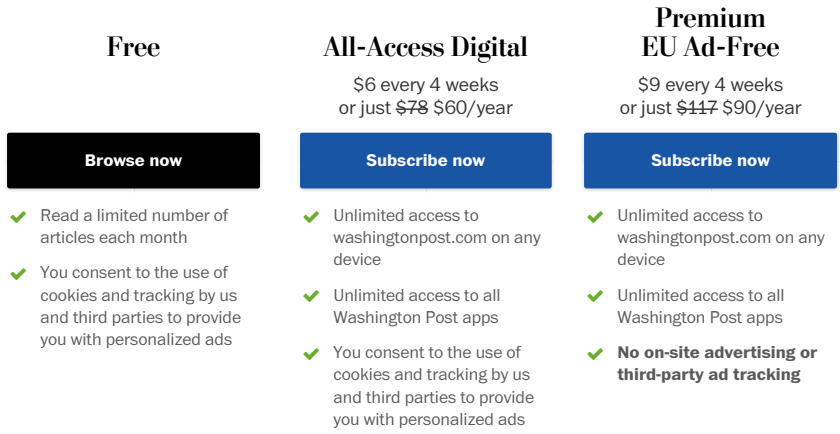


Fig. 6. The privacy notice displayed on the Washington Post website [141] requires users to pay to avoid tracking. Although this selection is unacceptable under the GDPR, enforcing it is problematic as well because The Washington Post is a US company.

Deletion Requests. In many cases, service providers are also limited regarding their capability to comply with the GDPR. Examples of these limitations include database backups [109], cloud computing [6], and blockchain technology [57, 158]. These cases present legitimate challenges for data holders, since users may have agreed that their data is collected and stored at one point, but the GDPR requires that this consent can be retroactively retracted. However, backups are usually heavily compressed and incremental [109] and, thus, cannot be easily accessed. Since these constraints lead to a situation where performing the required changes is computationally very expensive, such mechanisms hold some potential for abuse [109]. Approaches like the "just-in-time" dataset creation described by Debruyne et al. [30] could help data controllers to ensure that their datasets are GDPR-compliant at any given time. Especially if the computation is outsourced to cloud service providers, proper anonymization and encryption are important measures to ensure that the data controller does not pass on personal data to third parties without the user's consent [6].

Blockchain technology at its current state is largely contradictory to the GDPR because of the immutability of content that is captured using current state-of-the-art implementations [57]. However, it constitutes a good example where users may claim legitimate interest. Here, the interest of the collective to not provide means of editing or deleting data retroactively might outweigh the individual's interest, since the service could not be provided effectively anymore otherwise [109]. The trust in a blockchain structure would potentially decrease drastically if transactions and data on the blockchain could be changed retroactively to fully comply with the GDPR. Recent studies already highlight the issues of illicit content on a blockchain [97]. As a potential countermeasure, means to prune outdated and unneeded data from blockchains are actively researched [98].

Takeaway. *To date, uncertainties regarding the GDPR's scope on data processing practices remain, such as browser fingerprinting and blockchain technology. Future legislation may resolve these disagreements or trigger specific legal actions in which interest groups play a crucial role. Data controllers from non-European countries are also affected by the GDPR, although its enforcement in these cases poses a bigger challenge than it does for European parties.*

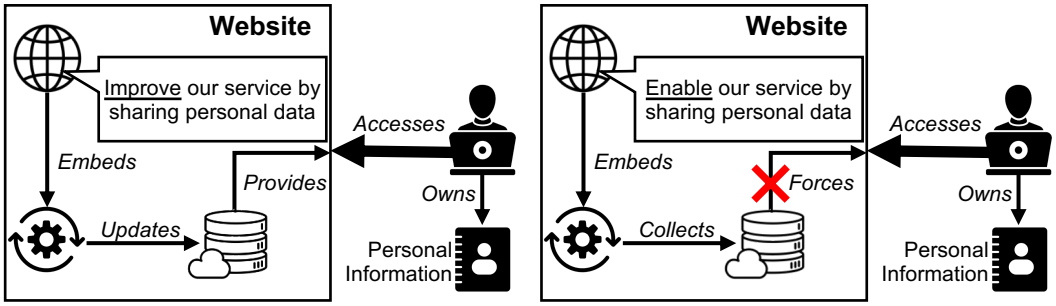
5.3 Data Portability

The right to data portability is one aspect of the GDPR, with plenty of speculation surrounding it [29, 69, 115]. Here, the main point of contention is the potential of sharing personal data among multiple platforms and services to oppose data monopolies, i.e., by preventing vendor lock-ins. The idea is that online platforms, such as social networks, e-commerce sites, and streaming services, share personal data of users among themselves on users' requests. This kind of data transfer could increase the users' control over their personal data, since they decide which services get access to which information and for what purpose [29, 69]. These models contradict the current situation, where parties often try to obtain as much data as possible regardless of the utility it provides for them in practice. Thus, data portability could also enable users to purposefully decentralize their personal information to prevent a single party from knowing more than they need to.

Competitive Advantage. However, we identify two major issues with such a line of thought. First, data holders may use the claim to legitimate interest to circumvent the otherwise mandatory transfer of users' personal information [69]. How such theoretical situations are handled in practice, and whether they are even realistic is up for debate, but parties are definitively incentivized to obtain as much personal information as possible and to share as little as possible to secure a competitive advantage based on the users' personal data [130]. Second, data holders can maximize their profit by expanding their services to provide as much personalization as possible, since this kind of service persuades individuals to grant web services access to their data. This situation might also introduce advantages for users. For example, Graef et al. [69] point out that this need for advances may increase competition and innovation, and impedes today's lock-ins. In contrast, the nature of some services is fundamentally at odds with personalization, such as public service media [130, 131]. The issue, which we see here, is that only being shown a curated selection of news stories carries the potential of skewing individuals' perception of the state of affairs, which may lead to or exacerbate social and political divides. Furthermore, public service institutions taking part in personal data mining may significantly erode trust in these institutions, creating contradictory goals [130, 131].

Market Consolidation. Even in the space of services that profit from personalization, we believe that such changes do not imply a rise in quality of experience, because currently, a handful of online service providers have no credible competitor in their respective domains. The advantage these companies have provides them with the necessary resources to expand their personalized functionalities much faster and wider than other service providers of the same domain [125]. With the possibility of data portability, large online companies may attempt to implement lock-outs to core features of their respective services, unless users agree to grant them access to their data. Taking a look at the market concentration of the online advertisement sector underlines this statement as it has demonstrably increased after the GDPR was enacted [21, 81]. We believe that the effect on data portability could be similar. In Figure 7, we illustrate two possible situations arising from the implementation of data portability.

Right Of Access. Simultaneously, data portability is also significantly impeded by how data controllers handle users' requests to obtain a copy of their data. Urban et al. [147] demonstrated that acquiring a copy of stored personal data from a data controller is commonly a multi-step process. In some cases, photocopies of ID cards have to be provided before data is made available [34, 147]. In most cases, user data is only shared after more than a week [147]. Boniface et al. [13] also found six of the largest tech companies (Google, Facebook, Microsoft, Instagram, Twitter, and LinkedIn) to handle personal data requests most securely and conveniently, which again may stand in the way of healthy competition. Although, as Graef et al. [69] point out, the process of sharing data between two data controllers does not have to be the same as how data is made available to users, the current



(a) This website encourages users to share their personal data. However, the service is also functional without such information.

(b) This website requires personal data sharing of its users. Otherwise, users are locked out from accessing the service.

Fig. 7. Two exemplary settings where users are either encouraged or required to share personal data. While enhancing a service with personal data on a voluntary basis to improve the user experience, implementing lock-outs unless data is shared is undesirable.

infrastructure to handle data requests requires improvements to allow for smoother transitions between similar services. Additional improvements are also needed to improve the security of data requests, as shown by Di Martino et al. [34]. Otherwise, requests by illegitimate parties might be handled by the data controller. In this context, blockchain technology may provide a solution to track the flow of personal information and manage user consent [107]. Kunz et al. [89] presented an approach of how personal data items and respective copies can be labeled, thus providing the required tracking functionality.

We can only speculate about future scenarios, as most web service providers have not yet implemented the necessary infrastructure to support the flexible exchange of personal data [29, 152]. Already in the past, web services usually supported several import formats (including imports from other platforms) and few to no ways to export any information from the respective platform. Furthermore, ongoing discussions on whether the right to data portability could also be interpreted to go along with data deletion, effectively migrating data from one holder to another [29, 115], are not concluded yet. Whether intended or not, such an interpretation could lead to user data becoming a rare and thus even more valuable commodity, which in turn increases the interest of data holders to lock-in users to their respective services. Although based on the specific GDPR provision, data portability does not have to be accompanied by erasure [69], users might prefer this interpretation to satisfy their interest in protecting their privacy.

Takeaway. *Considering the current state of the web ecosystem and its reliance on personalization, and thus personal data, we are skeptical that data portability has an overall positive effect on privacy. Furthermore, we find little evidence that data controllers spend a lot of effort into implementing the required interfaces for portability.*

5.4 Privacy by Default

Although the GDPR holds the potential to improve the user experience of web services, in the long run [29], some of the immediate consequences of its enactment were less positive. On the one hand, a number of websites disabled access from EU states in May of 2018 [25, 72], which could indicate them feeling vulnerable on the basis of the GDPR, as they had not implemented the necessary changes to their data policies. On the other hand, leading online companies, such as Amazon, Facebook, and Google, seem to have started investing in compliance with the new EU regulation

early on [81, 93]. On these platforms, the number of trackers linked to these large companies has declined measurably, with Facebook being the most affected and Google the least [93], indicating an effort to comply with the GDPR.

Lobbyism. At the same time, these companies spend large sums of money lobbying their interests to US lawmakers, spending a combined amount of roughly 45 million $ on matters regarding privacy in 2018 alone [32]. However, the interest of these companies to collect data is not limited to selling personal information, since they also use this data to improve the personalization and performance of their services and conduct research [53, 68]. Thus, many of these parties' business models are directly linked to their ability to collect data.

Technological Impact. Tene et al. [137] point out that the GDPR's call for privacy by default [51, Article 25] shifts the challenge of privacy from a question of law and policies to a matter of technology. Arguments, whether certain practices fall under legitimate interest or asking users to consent to a myriad of trackers, could thus be rendered obsolete by implementing privacy-preserving technologies for data processing. Concepts such as Bloom filters [11], differential privacy [35], oblivious transfer [59], and homomorphic encryption [1] have been proposed before the GDPR's enactment and might constitute valuable tools to achieve this goal. Newer approaches draw on these techniques to implement efficient privacy-preserving protocols [43, 101], and industry leaders are adopting them for their own use [85].

This situation might further contribute to the divide between smaller service providers, which have been more negatively affected by the GDPR, and these larger companies, which are increasingly implementing privacy-preserving measures [21, 81, 137]. How this situation evolves in the future is difficult to predict due to the complexity of the matter, but the market for privacy-enhancing technology is definitively increasing [75, 122], and technology-focused start-ups with an emphasis on privacy seemingly profit from the GDPR [122, 137].

Takeaway. *The GDPR features numerous provisions that may have contributed to an observable shift in companies' data collection policies, most notably Article 25, which calls for privacy by default. Apart from users, this change seems to benefit both large companies implementing privacy-preserving measures as well as privacy-focused start-ups.*

6 GDPR-RELEVANT TAKEAWAYS FOR SERVICE PROVIDERS ON THE WEB

In the previous sections, we provided a broad discussion of the GDPR's impact on web-based services and discussed relevant implications. To wrap up our discussion, we derived four guidelines for service providers that are mainly inspired by suggestions and arguments raised by authors of the investigated literature. The guidelines should not be regarded as legal advice. Instead, they serve to outline a shift towards a more user-centric design of online services that is beneficial to both the users and providers of these services. For simplicity, we consider service providers, whom we assume are interested in achieving GDPR-compliance, without having to implement more changes to their data processing practices than strictly required. In the interest of presenting a balanced discussion, we also consider the end-user's position, whom we assume to be mainly interested in data minimization and convenience.

Communicating with Customers. Based on the data we have analyzed, improving the communication on the part of service providers may, in many cases, have been sufficient to avoid legal actions and fines. In most cases, users were either unaware of certain practices of service providers or explicitly disagreed with them, and thus, no informed consent was given [7, 26, 120]. Due to our computer science background, we cannot rate whether such practices would have been within the boundaries of the law otherwise. However, notably, the main issue with fulfilling the GDPR seems to be inaction and carelessness when handling personal data, exemplified by the low priority many

businesses seem to place on the role of the data protection officer [33, 104]. For service providers, who strive to comply with the GDPR, the data protection officer is an important asset, since they can also help to clear up uncertainties regarding the legal conditions [6]. Data controllers should take into account that the GDPR also adds a considerable workload through requests under the right of access [34, 147] and the need to update the existing backup management [109]. Boniface et al. [13] suggest using cryptographic hash functions, which enable secure and anonymous authentication and prove ownership of personal information items. The proposed system would render the problematic practices, such as requiring additional personal information, e.g., copies of ID cards, obsolete. However, the system proposed by Boniface et al. [13] only works if cookies are created client-side. Thus, implementing the necessary routines to comply with user requests may take time and consume resources, especially for smaller enterprises.

Certainly, the GDPR accounts for these challenges, allowing deadlines to such requests to be extended in cases where service providers are overwhelmed [51, Article 12 (3)]. Furthermore, Article 30 exempts institutions with less than 250 employees from some of the privacy-related duties, and the GDPR specifically states that "[...] supervisory authorities [...] are encouraged to take account of the specific needs of [...] small and medium-sized enterprises [...]" [51, Recital 13]. Thus, we want to stress that ignoring user requests is not a suitable solution to this problem under any circumstance, as this behavior hurts the users' trust even further. Regarding our delivery example from Section 3.2.1, a pizza delivery can be expected to store one's address to carry out a delivery. However, retaining this information with the purpose of advertisement is not covered by the concept of legitimate interest, as demonstrated by the case of Delivery Hero [26]. Consequentially, all relevant information must be deleted upon the customer's request in a timely manner. We found multiple proposals and experimental implementations of blockchain-based personal data management systems that automate such processes [95, 107]. Blockchain technology is especially useful in this regard, since it provides decentralized consensus and authentication. Thus, data subjects can trace how and by whom their personal information is processed, and data controllers do not have to implement complicated authentication procedures to confirm data subjects' identity. Mahindrakar and Joshi [95] even claim that their system is (technically) capable of fulfilling the tasks of the data protection officer, although this may be legally problematic. Whether the implementation and maintenance of such a (rather complex) system, in the end, reduces the workload of data controllers depends on many factors. We would nevertheless recommend that online service providers seriously evaluate the cost-benefit trade-off, the mentioned systems provide.

Takeaway. *We identify effective communication with users as a central requirement to prevent legal repercussions in which the data protection officer plays a currently undervalued role. Related work outlines GDPR-compliant methods how processes, such as deletion or access requests, can be automated, which we would advise data controllers to take into consideration.*

Provide Transparency. Based on observations pertaining to the amount of time users have to spend on reading privacy policies in theory [99] and the time they really spend [71, 138], it is clear that the primary purpose of today's privacy policies is legal compliance. Web service providers are aware of the discrepancy between what is legally required and what provides a practical benefit to the user [146], and informing users that they are being tracked is linked to an increased risk perception [16]. Thus, we understand that service providers may feel as if informing users regarding their employed tracking practices beyond what is minimally legally required is not in their interest [123]. However, undisclosed cookie usage has been shown to significantly reduce user trust, while disclosed cookie usage has an almost negligible effect [102]. Therefore, we endorse user-centric systems, such as *Privee* [160], *PrivacyGuide* [139], and *Polisis* [55], which may benefit service providers as much as their users. Furthermore, the utility of the named systems extends

beyond informing users to aid both service providers and data protection authorities in determining whether a given privacy policy is GDPR-conform [138, 139]. Considering that some data controllers call for more specific instructions by data protection authorities [146] and privacy researchers advocate for stronger involvement of data protection authorities [16, 105, 145], these tools help to clarify and enforce the GDPR in online settings, improving privacy standards.

Takeaway. *Grading systems developed by researchers are beneficial to users by clarifying privacy risks, to data protection authorities by identifying potential breaches of privacy legislation, and service providers by increasing consumer trust and warning them of illegitimate data processing policies.*

Enable Privacy Preferences. We additionally endorse the implementation of conservative cookie policies that refrain from setting cookies right away, and providing users with at least the amount of control of a Type 3 cookie banner (cf. Figure 2). This policy also limits the amount of cleanup that applications have to perform, once a user does not give or retracts consent, which is especially problematic when also implementing third-party trackers. In these settings, consent management platforms may provide the least-effort solution for data controllers to implement the level of control required by the GDPR, but are not without issues. First, CMP banners still require some effort to configure properly [31, 105]. Second, although they provide better control to users than most other solutions [31, 73], CMP banners commonly do not provide a convenient way to avoid tracking. The main problems in this regard are (needlessly) complicated interface design [73], misleading cookie classifications [31], and visual biasing against privacy-preserving settings [105, 128]. All of these factors contribute to users' frustrations with cookie banners [149]. Thus, we conclude that this situation does not provide a satisfying solution for data subjects.

We instead suggest web service hosts to consider the adoption of systems, as proposed by Gerl et al. [63, 64], which automatically negotiate privacy preferences upon loading a website. The proposed system is based on the Layered Privacy Language [62] and would allow users to specify granular and universal privacy preferences. The website then either respects the set preferences, or informs the user about conflicting settings, who can, in turn, decide how to proceed. The entire privacy negotiation process would happen before any data is processed and thus, guarantee both protection of the data subject's information as well as automated compliance with the GDPR. We understand that data controllers have an interest in processing data that can be legitimate in nature. If this processing is happening against an individual's explicitly expressed wishes or without their knowledge, however, it is no longer legitimate [156], and resulting fines may easily outweigh the profits obtained through illicit data processing. Furthermore, giving users a choice to opt out of tracking may increase their satisfaction with a given website [123]. Since users are interested in both privacy and convenient browsing [149], which is enabled by automated privacy preference negotiation [64], it is also in the interest of data controllers to provide the necessary amount of control without unnecessary obstructions. Until service providers broadly adopt such approaches, we would advise users of the respective services to look into privacy-preserving means, such as privacy orientated web browsers, search engines, and browser extensions which have been shown to overall limit tracking [42] without negatively affecting performance [15].

Takeaway. *Since the line between what is acceptable and what is not under the GDPR has not yet been firmly established, we recommend that users are provided with a fair and equal choice without tricking users or annoying them with endless submenus. The GDPR specifies these requirements, and improper use may result in significant fines. In contrast to these fines, providing a convenient and privacy-preserving service may increase its revenue, since it matches users' priorities.*

Avoid Storing Raw Data. Finally, services may gain an advantageous position by processing all personal information they gain access to in a transformational fashion, because such practices support their claim to legitimate interest, which is a potentially powerful defense for service

providers to avoid erasure of collected data [29, 69]. If the aggregation process, for example, discards all information other than non-personal metadata, this data is no longer protected under the GDPR [69]. Data processed in such a manner can then also be shared with third parties either to integrate an external service, such as advertisement or service optimization, or to outsource computations [6]. Thus, user data should be collected and processed with a specific purpose in mind, which is communicated to the user and does not leak personal information to any third party.

Purely acquiring large amounts of raw data, possibly with an intention to sell it, is at odds with the GDPR (cf. Section 4.4). Encryption on its own is unlikely to pass as a sufficient claim to legitimate interest, but definitively recommended as it reduces the risk of a data breach of personal data, resulting in a loss of users' trust as well as fines under the GDPR [116, 121]. The principle of data minimization is mentioned in multiple parts of the GDPR. Thus, it provides a good standard for service providers if they want to achieve compliance with this new regulation. Again, existing research [30, 158] may aid data controllers in finding ways to implement GDPR-compliant data processing pipelines. Furthermore, market data shows that privacy-focused companies profit from the recent privacy legislation [75, 122, 137] creating a financial incentive to use privacy-preserving technologies.

Takeaway. *Instead of spending resources on finding and implementing legal workarounds, privacy by default presents a much more stable approach, in our opinion. As new privacy-preserving technologies emerge, companies that invest in such functionality can profit early on by offering high-privacy standards to their users, both an attractive property from a user perspective and a reduction of regulatory overhead.*

7 CONCLUSION

In this paper, we presented a literature-based analysis of the impact the GDPR has had on users and service providers with a focus on the World Wide Web. We summarized results obtained by studies that collected and evaluated quantitative measures related to the GDPR (cf. Section 4.2) with the goal of giving the reader an overview of the landscape before and after its enactment. Thereby, we conclude that the GDPR has had an overall positive effect on privacy on the web by increasing transparency and user control w.r.t. tracking, although they fall short of the strict requirements set by the GDPR in most cases. Based on our broad review of multiple scientific sources, we identified a dependency on third-party content mixed with uncertainties regarding the regulatory framework as the central currently unresolved issue.

Unfortunately, the amount of available data related to the impact of the GDPR on the web is still rather limited. Through our investigation, we highlighted measures that could provide valuable insight into the current situation on the web regarding personal privacy. Based on existing research, we also give some suggestions on how studies could be designed to obtain such data (cf. Section 5.1), for example, by specifically repeating study designs from before the GDPR and comparing the results. Differences in the results may point towards certain shifts in tracking technology.

Takeaways. Our main takeaways regarding the implementation of GDPR rules are as follows. Adjusting services constitutes a continual process where many platforms are still several steps away from achieving conformity. Although the GDPR significantly contributed to an EU-wide alignment of privacy levels which previous legislation failed to accomplish, the high-privacy standards set by the GDPR are only fulfilled by the minority of web services.

From our perspective, users also have a major part in this process, since they are the ones who can claim the right to data autonomy, and through that create incentives for service providers to extend their privacy-enhancing technologies without sacrificing usability. As exemplified by findings pertaining to third-party tracking and the effectiveness of cookie banners, users cannot trust that

the GDPR on its own compels service providers to limit their data collection and processing. The GDPR should rather be viewed as a tool, individuals can take advantage of to protect their privacy.

Furthermore, we highlight that service providers are under a considerable amount of pressure to adapt their policies to comply with the new regulation and remain competitive in the changing online ecosystem. In this regard, we observe some speculation over the changing flow of personal data between service providers. We are unable to provide a conclusive statement on how individuals will be enabled to take advantage of GDPR rules in tomorrow's online ecosystem, but remain critical whether users are truly and fully in control of data that concerns them personally.

Future Work. Based on the material we have analyzed over the course of this paper, we find that tracking and advertisement on the World Wide Web are currently undergoing some changes. The market concentration of tracking services has increased, and advertisement providers are adapting their data processing technology. Once the ePrivacy Regulation (cf. Section 3.2.3) is passed and enters into force, data processing on the web might once again have to undergo considerable changes. Thus, with its enactment, more studies are needed to yet again capture the respective changes and trends. While we mostly focused on the aspect of privacy, the economic impact of the GDPR is just as important, especially if we assume that revenue is one of or the most important evaluation criteria for a service provider to assess the performance of their service. Unfortunately, our corpus did not cover many scientific works related to the economy. Thus, we or others will have to investigate this area in a future report.

Furthermore, we limited ourselves to the impact of the GDPR on the web, because the web is an important aspect of many people's everyday life. However, our method also returned numerous scientific works related to the Internet of Things and medicine, both of which constitute their own fascinating areas of research. We leave the analysis of the GDPR's overall impact on these topics for future research as well.

Lastly, we identified the right to data portability as an intriguing research subject, due the profound effect on data processing it could have on the web and society at large. From our point of view, the required infrastructure, and services that would take advantage of the feature to allow users to specifically share personal data, are currently non-existent. Still, the GDPR mandates services to make personal data they store available to their users, and thus privacy management platforms that incorporate interfaces to request and manage personal information may be put into practice fairly soon. Again, our corpus of research did not feature many works on that topic, so we will investigate data portability in future work.

ACKNOWLEDGMENT

The first author would like to acknowledge the invaluable support of his co-authors that has been crucial to the progress of this work, and vows to devote more time on chores again, once it has been published.

REFERENCES

[1] Abbas Acar, Hidayet Aksu, A Selcuk Uluagac, and Mauro Conti. 2018. A Survey on Homomorphic Encryption Schemes: Theory and Implementation. *Comput. Surveys* 51, 4 (2018), 1–35. https://doi.org/10.1145/3214303

[2] Gunes Acar, Marc Juarez, Nick Nikiforakis, Claudia Diaz, Seda Gürses, Frank Piessens, and Bart Preneel. 2013. FPDetective: Dusting the Web for Fingerprinters. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security (CCS '13)*. ACM, 1129–1140. https://doi.org/10.1145/2508859.2516674

[3] Nasser Mohammed Al-Fannah, Wanpeng Li, and Chris J. Mitchell. 2018. Beyond Cookie Monster Amnesia: Real World Persistent Online Tracking. In *Proceedings of the 21st International Conference on Information Security (ISC '18)*, Vol. 11060. Springer, 481–501. https://doi.org/10.1007/978-3-319-99136-8_26

[4] Jan Philipp Albrecht. 2016. How the GDPR will Change the World. *European Data Protection Law Review* 2, 3 (2016), 287–289. https://doi.org/10.21552/EDPL/2016/3/4

[5] Alexa Internet, Inc. 1996. Alexa - Keyword Research, Competitive Analysis, & Website Ranking. https://www.alexa.com/.

[6] Alaa Altorbaq, Fredrik Blix, and Stina Sörman. 2017. Data subject rights in the cloud: A grounded study on data protection assurance in the light of GDPR. In *Proceedings of the 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST '12)*. IEEE, 305–310. https://doi.org/10.23919/ICITST.2017.8356406

[7] Heike Anger and Dietmar Neuerer. 2019 (accessed April 12, 2020). DSGVO in Zahlen: Hoher Aufwand, aber auch Ertrag. https://www.handelsblatt.com/politik/deutschland/bussgeld-bilanz-dsgvo-in-zahlen-hoher-aufwand-aber-auch-ertrag/24361018.html.

[8] APA-OTS. 2019 (accessed April 12, 2020). Strafverfahren gegen Österreichische Post AG. https://www.ots.at/presseaussendung/OTS_20191029_OTS0095/strafverfahren-gegen-oesterreichische-post-ag.

[9] Paul Ashley, Satoshi Hada, Günter Karjoth, Calvin Powers, and Matthias Schunter. 2003. Enterprise Privacy Authorization Language (EPAL 1.2). W3C Member Submission SUBM-EPAL-20031110.

[10] Autoriteit Persoonsgegevens. 2019 (accessed June 28, 2020). Websites moeten toegankelijk blijven bij weigeren tracking cookies. https://autoriteitpersoonsgegevens.nl/nl/nieuws/websites-moeten-toegankelijk-blijven-bij-weigeren-tracking-cookies.

[11] Burton H. Bloom. 1970. Space/Time Trade-Offs in Hash Coding with Allowable Errors. *Commun. ACM* 13, 7 (1970), 422–426. https://doi.org/10.1145/362686.362692

[12] Károly Boda, Ádám Máté Földes, Gábor György Gulyás, and Sándor Imre. 2011. User Tracking on the Web via Cross-Browser Fingerprinting. In *Proceedings of the 16th Nordic Conference on Secure IT Systems (NordSec '11)*, Vol. 7161. Springer, 31–46. https://doi.org/10.1007/978-3-642-29615-4_4

[13] Coline Boniface, Imane Fouad, Nataliia Bielova, Cédric Lauradoux, and Cristiana Santos. 2019. Security Analysis of Subject Access Request Procedures. In *Proceedings of the 7th Annual Privacy Forum on Privacy Technologies and Policy (APF '19)*, Vol. 11498. Springer, 182–209. https://doi.org/10.1007/978-3-030-21752-5_12

[14] Emma Bordessa. 2018 (accessed April 12, 2020). ICO statistics show increase in reported incidents ahead of GDPR. https://www.itgovernance.co.uk/blog/ico-statistics-show-increase-in-reported-incidents-ahead-of-gdpr.

[15] Kevin Borgolte and Nick Feamster. 2020. Understanding the Performance Costs and Benefits of Privacy-focused Browser Extensions. In *Proceedings of The Web Conference 2020 (WWW '20)*. IW3C2, 2275–2286. https://doi.org/10.1145/3366423.3380292

[16] Rico Bornschein, Lennard Schmidt, and Erik Maier. 2020. The Effect of Consumers' Perceived Power and Risk in Digital Information Privacy: The Example of Cookie Notices. *Journal of Public Policy & Marketing* 39, 2 (2020), 135–154. https://doi.org/10.1177/0743915620902143

[17] Luca Bufalieri, Massimo La Morgia, Alessandro Mei, and Julinda Stefa. 2020. GDPR: When the Right to Access Personal Data Becomes a Threat. In *Proceedings of the 2020 IEEE International Conference on Web Services (ICWS '20)*. IEEE, 75–83. https://doi.org/10.1109/ICWS49710.2020.00017

[18] Eric Butler, John Teddy, and Martin Waugh. 2006. First-party cookie for tracking web traffic. Patent US20060265495A1.

[19] Eric Butler, John Teddy, and Martin Waugh. 2012. Method for cross-domain tracking of web site traffic. Patent US8131861B2.

[20] José González Cabañas, Ángel Cuevas, and Rubén Cuevas. 2018. Unveiling and Quantifying Facebook Exploitation of Sensitive Personal Data for Advertising Purposes. In *Proceedings of the 27th USENIX Conference on Security Symposium (SEC '18)*. USENIX Association, 479–495.

[21] Cliqz GmbH. 2018 (accessed April 12, 2020). GDPR - What happened? https://whotracks.me/blog/gdpr-what-happened.html.

[22] CMS Legal. 2019 (accessed May 16, 2020). e-Privacy. https://cms.law/en/deu/insight/e-privacy.

[23] Court of Justice of the European Union. 2019 (accessed June 28, 2020). Storing cookies requires internet users' active consent. https://curia.europa.eu/jcms/upload/docs/application/pdf/2019-10/cp190125en.pdf.

[24] Cybot. 2020 (accessed June 28, 2020). Cookie walls | EDPB guidelines on cookie walls and valid consent. https://www.cookiebot.com/en/cookie-walls/.

[25] Adrian Dabrowski, Georg Merzdovnik, Johanna Ullrich, Gerald Sendera, and Edgar Weippl. 2019. Measuring cookies and web privacy in a post-gdpr world. In *Proceedings of the 20th International Conference on Passive and Active Network Measurement (PAM '19)*, Vol. 11419. Springer, 258–270. https://doi.org/10.1007/978-3-030-15986-3_17

[26] Ingo Dachwitz. 2019 (accessed April 12, 2020). Berliner Datenschutzbehörde verhängt bisher höchstes DSGVO-Bußgeld gegen Lieferdienst. https://netzpolitik.org/2019/berliner-datenschutzbehoerde-verhaengt-bisher-hoechstes-dsgvo.

[27] Brian Daigle and Mahnaz Khan. 2019 (accessed June 21, 2020). One Year In: GDPR Fines and Investigations against U.S.-Based Firms. https://www.usitc.gov/publications/332/executive_briefings/gdpr_enforcement.pdf.

[28] Ben Davis. 2017 (accessed April 12, 2020). GDPR for marketers: Five examples of 'Legitimate Interests'. https://econsultancy.com/gdpr-for-marketers-five-examples-of-legitimate-interests/.

[29] Paul De Hert, Vagelis Papakonstantinou, Gianclaudio Malgieri, Laurent Beslay, and Ignacio Sanchez. 2018. The right to data portability in the GDPR: Towards user-centric interoperability of digital services. *Computer Law & Security Review* 34, 2 (2018), 193–203. https://doi.org/10.1016/j.clsr.2017.10.003

[30] Christophe Debruyne, Harshvardhan J. Pandit, Dave Lewis, and Declan O'Sullivan. 2020. "Just-in-time" generation of datasets by considering structured representations of given consent for GDPR compliance. *Knowledge and Information Systems* (2020). https://doi.org/10.1007/s10115-020-01468-x

[31] Martin Degeling, Christine Utz, Christopher 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. *Proceedins of the 26th Annual Network and Distributed System Security Symposium (NDSS '19)* (2019). https://doi.org/10.14722/ndss.2019.23378

[32] A. J. Dellinger. 2019 (accessed June 21, 2020). How The Biggest Tech Companies Spent Half A Billion Dollars Lobbying Congress. https://www.forbes.com/sites/ajdellinger/2019/04/30/how-the-biggest-tech-companies-spent-half-a-billion-dollars-lobbying-congress/.

[33] Der Landesbeauftragte für den Datenschutz und die Informationsfreiheit Baden-Württemberg. 2019 (accessed June 14, 2020). Evaluierung der DS-GVO - Ergebnisse der Anhörung des LfDI BW zur Evaluierung der DS-GVO. https://www.baden-wuerttemberg.datenschutz.de/wp-content/uploads/2019/12/Evalierungsbericht-Stand-17.12.2019.pdf.

[34] Mariano Di Martino, Pieter Robyns, Winnie Weyts, Peter Quax, Wim 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)*. USENIX Association, 371–386.

[35] Cynthia Dwork. 2006. Differential Privacy. In *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming (ICALP '06)*, Vol. 4052. Springer, 1–12. https://doi.org/10.1007/11787006_1

[36] Peter Eckersley. 2009 (accessed April 12, 2020). How Online Tracking Companies Know Most of What You Do Online (and What Social Networks Are Doing to Help Them). https://www.eff.org/deeplinks/2009/09/online-trackers-and-social-networks.

[37] Peter Eckersley. 2010 (accessed April 12, 2020). Browser Versions Carry 10.5 Bits of Identifying Information on Average. https://www.eff.org/deeplinks/2010/01/tracking-by-user-agent.

[38] Jens Eckhardt and Rudi Kramer. 2013. EU-DSGVO—Diskussionspunkte aus der Praxis. *Datenschutz und Datensicherheit - DuD* 37, 5 (2013), 287–294. https://doi.org/10.1007/s11623-013-0110-5

[39] Electronic Frontier Foundation. 2011 (accessed April 12, 2020). Tracking Users: From Cookie to Device Fingerprinting. https://www.eff.org/issues/do-not-track.

[40] Electronic Frontier Foundation. 2020. Defending your rights in the digital world. https://www.eff.org/.

[41] Electronic Privacy Information Center. 2000. *Pretty Poor Privacy: An Assessment of P3P and Internet Privacy.* Technical Report. Electronic Privacy Information Center. https://epic.org/reports/prettypoorprivacy.html.

[42] Steven Englehardt and Arvind Narayanan. 2016. Online Tracking: A 1-million-site Measurement and Analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16)*. ACM, 1388–1401. https://doi.org/10.1145/2976749.2978313

[43] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. 2014. RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS '14)*. ACM, 1054–1067. https://doi.org/10.1145/2660267.2660348

[44] eTorch Inc. 2018 (accessed April 12, 2020). Legitimate interest. https://www.gdpreu.org/the-regulation/key-concepts/legitimate-interest/.

[45] European Commission. 2017 (accessed May 16, 2020). Proposal for an ePrivacy Regulation. https://ec.europa.eu/digital-single-market/en/proposal-eprivacy-regulation.

[46] European Commission. 2018 (accessed March 16, 2021). What does 'grounds of legitimate interest' mean? https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/grounds-processing/what-does-grounds-legitimate-interest-mean_en.

[47] European Parliament. 2017. First Reading Report on the proposal for a regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications). COM(2017)0010 – C8-0009/2017 – 2017/0003(COD).

[48] European Parliament and Council. 1995. Data Protection Directive. Directive 95/46/EC.

[49] European Parliament and Council. 2002. Privacy and Electronic Communications Directive. Directive 2002/58/EC.

[50] European Parliament and Council. 2009. Amendment of Privacy and Electronic Communications Directive. Directive 2009/136/EC.

[51] European Parliament and Council. 2016. General Data Protection Regulation. Regulation (EU) 2016/679.

[52] European Union. 2016 (accessed April 12, 2020). Regulations, Directives and other acts. https://europa.eu/european-union/eu-law/legal-acts_en.

[53] Facebook Inc. 2018 (accessed June 22, 2020). Data Policy. https://www.facebook.com/about/privacy.
[54] Amin FaizKhademi, Mohammad Zulkernine, and Komminist Weldemariam. 2015. FPGuard: Detection and Prevention of Browser Fingerprinting. In *Proceedings of the 29th Annual IFIP WG 11.3 Working Conference on Data and Applications Security and Privacy XXIX (DBSec '15)*, Vol. 9149. Springer, 293–308. https://doi.org/10.1007/978-3-319-20810-7_21
[55] Kassem Fawaz, Thomas Linden, and Hamza Harkous. 2019. Invited Paper: The Applications of Machine Learning in Privacy Notice and Choice. In *Proceedings of the 2019 11th International Conference on Communication Systems Networks (COMSNETS '19)*. IEEE, 118–124. https://doi.org/10.1109/COMSNETS.2019.8711280
[56] Alison Fennah and Marie-Clare Puffett. 2018 (accessed May 16, 2020). IAB Europe research: AdEx Benchmark 2017 Study. https://iabeurope.eu/research-thought-leadership/iab-europe-report-adex-benchmark-2017-report/.
[57] Valeria Ferrari, João Pedro Quintais, Alexandra Giannopoulou, and Balázs Bodó. 2018. EU Blockchain Observatory and Forum Workshop on GDPR, Data Policy and Compliance. Institute for Information Law Research Paper No. 2018-04. https://doi.org/10.2139/ssrn.3247494
[58] Roy Fielding, Jim Gettys, Jeffrey Mogul, Henrik Frystyk, Larry Masinter, Paul Leach, and Tim Berners-Lee. 1999. Hypertext Transfer Protocol – HTTP/1.1. IETF RFC 2616.
[59] Michael J. Fischer, Silvio Micali, and Charles Rackoff. 1996. A secure protocol for the oblivious transfer. *Journal of Cryptology* 9, 3 (1996), 191–195. https://doi.org/10.1007/BF00208002
[60] Matus Formanek, Erika Sustekova, and Vladimir Filip. 2019. The progress of web security level related to European open access LIS repositories between 2016 and 2018. *JLIS. it* 10, 2 (2019), 107–115. https://doi.org/10.4403/jlis.it-12545
[61] Barton Gellman and Laura Poitras. 2013 (accessed June 21, 2020). U.S., British intelligence mining data from nine U.S. Internet companies in broad secret program. https://www.washingtonpost.com/investigations/us-intelligence-mining-data-from-nine-us-internet-companies-in-broad-secret-program/2013/06/06/3a0c0da8-cebf-11e2-8845-d970ccb04497_story.html.
[62] Armin Gerl, Nadia Bennani, Harald Kosch, and Lionel Brunie. 2018. *LPL, Towards a GDPR-Compliant Privacy Language: Formal Definition and Usage*. Springer, 41–80. https://doi.org/10.1007/978-3-662-57932-9_2
[63] Armin Gerl and Bianca Meier. 2019. The Layered Privacy Language Art. 12–14 GDPR Extension–Privacy Enhancing User Interfaces. *Datenschutz und Datensicherheit - DuD* 43, 12 (2019), 747–752. https://doi.org/10.1007/s11623-019-1200-9
[64] Armin Gerl, Bianca Meier, and Stefan Becher. 2019. Let Users Control Their Data–Privacy Policy-Based User Interface Design. In *Proceedings of the 1st International Conference on Human Interaction and Emerging Technologies (IHIET '19)*, Vol. 1018. Springer, 790–795. https://doi.org/10.1007/978-3-030-25629-6_123
[65] Ben Glanville. 2018 (accessed April 18, 2020). 72% of Brits haven't heard about GDPR. https://yougov.co.uk/topics/politics/articles-reports/2018/03/01/72-brits-havent-heard-about-gdpr.
[66] Samuel Goldberg, Garrett Johnson, and Scott Shriver. 2019. Regulating Privacy Online: The Early Impact of the GDPR on European Web Traffic & E-Commerce Outcomes. SSRN. https://doi.org/10.2139/ssrn.3421731
[67] Google Developers. 2016 (accessed April 12, 2020). User Opt-out. https://developers.google.com/analytics/devguides/collection/analyticsjs/user-opt-out.
[68] Google Inc. 2020 (accessed June 22, 2020). Google Privacy Policy. https://policies.google.com/privacy?hl=en.
[69] Inge Graef, Martin Husovec, and Nadezhda Purtova. 2018. Data Portability and Data Control: Lessons for an Emerging Concept in EU Law. *German Law Journal* 19, 6 (2018), 1359–1398. https://doi.org/10.1017/S2071832200023075
[70] Samuel Greengard. 2018. Weighing the impact of GDPR. *Commun. ACM* 61, 11 (2018), 16–18. https://doi.org/10.1145/3276744
[71] Rüdiger Grimm and Alexander Rossnagel. 2000. Can P3P help to protect privacy worldwide?. In *Proceedings of the 2000 ACM Workshops on Multimedia (MULTIMEDIA '00)*. ACM, 157–160. https://doi.org/10.1145/357744.357917
[72] Alex Hern and Jim Waterson. 2018 (accessed April 12, 2020). Sites block users, shut down activities and flood inboxes as GDPR rules loom. https://www.theguardian.com/technology/2018/may/24/sites-block-eu-users-before-gdpr-takes-effect.
[73] Maximilian Hils, Daniel W. Woods, and Rainer Böhme. 2020. Measuring the Emergence of Consent Management on the Web. In *Proceedings of the ACM Internet Measurement Conference (IMC '20)*. ACM, 317–332. https://doi.org/10.1145/3419394.3423647
[74] Xuehui Hu and Nishanth Sastry. 2019. Characterising Third Party Cookie Usage in the EU after GDPR. In *Proceedings of the 10th ACM Conference on Web Science (WebSci '19)*. ACM, 137–141. https://doi.org/10.1145/3292522.3326039
[75] Victoria Hudgins. 2019 (accessed April 12, 2020). Data Privacy Market Still Has Room for All Entrants. https://www.law.com/legaltechnews/2019/07/18/data-privacy-market-still-has-room-for-all-entrants/.
[76] Information Commissioner's Office. 1984. ICO. https://ico.org.uk/.
[77] Costas Iordanou, Georgios Smaragdakis, Ingmar Poese, and Nikolaos Laoutaris. 2018. Tracing Cross Border Web Tracking. In *Proceedings of the 2018 ACM Internet Measurement Conference (IMC '18)*. ACM, 329–342. https://doi.org/10.1145/3278532.3278561

[78] Timo Jakobi, Gunnar Stevens, Anna-Magdalena Seufert, Max Becker, and Max von Grafenstein. 2020. Web Tracking Under the New Data Protection Law: Design Potentials at the Intersection of Jurisprudence and HCI. *i-com* 19, 1 (2020), 31–45. https://doi.org/10.1515/icom-2020-0004

[79] Yousra Javed, Khondaker Musfakus Salehin, and Mohamed Shehab. 2020. A Study of South Asian Websites on Privacy Compliance. *IEEE Access* 8 (2020), 156067–156083. https://doi.org/10.1109/ACCESS.2020.3019334

[80] Carlos Jensen and Colin Potts. 2004. Privacy Policies as Decision-Making Tools: An Evaluation of Online Privacy Notices. In *Proceedings of the 2004 ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. ACM, 471–478. https://doi.org/10.1145/985692.985752

[81] Garrett Johnson and Scott Shriver. 2019. Privacy & Market Concentration: Intended & Unintended Consequences of the GDPR. SSRN. https://doi.org/10.2139/ssrn.3477686

[82] Judgment of the Court (Grand Chamber). 2014. Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González. C-131/12 - Google Spain and Google.

[83] Chinmay Kakatkar and Martin Spann. 2019. Marketing analytics using anonymized and fragmented tracking data. *International Journal of Research in Marketing* 36, 1 (2019), 117–136. https://doi.org/10.1016/j.ijresmar.2018.10.001

[84] Daphne Keller. 2015 (accessed April 12, 2020). The Final Draft of Europe's "Right to Be Forgotten" Law. http://cyberlaw.stanford.edu/blog/2015/12/final-draft-europes-right-be-forgotten-law.

[85] Krishnaram Kenthapadi, Ilya Mironov, and Abhradeep Guha Thakurta. 2019. Privacy-preserving Data Mining in Industry. In *Proceedings of the 12th ACM International Conference on Web Search and Data Mining (WSDM '19)*. ACM, 840–841. https://doi.org/10.1145/3289600.3291384

[86] Richie Koch. 2019 (accessed April 12, 2020). Cookies, the GDPR, and the ePrivacy Directive. https://gdpr.eu/cookies/.

[87] Alžběta Krausová. 2018. Online Behavior Recognition: Can We Consider It Biometric Data under GDPR? *Masaryk University Journal of Law and Technology* 12, 2 (2018), 161–177.

[88] Oksana Kulyk, Annika Hilt, Nina Gerber, and Melanie Volkamer. 2018. "This Website Uses Cookies": Users' Perceptions and Reactions to the Cookie Disclaimer. In *Proceedings of the 3rd European Workshop on Usable Security (EuroUSEC '18)*. Internet Society. https://doi.org/10.14722/eurousec.2018.23012

[89] Immanuel Kunz, Valentina Casola, Angelika Schneider, Christian Banse, and Julian Schütte. 2020. Towards Tracking Data Flows in Cloud Architectures. In *Proceedings of the 2020 IEEE 13th International Conference on Cloud Computing (CLOUD '20)*. IEEE, 445–452. https://doi.org/10.1109/CLOUD49709.2020.00066

[90] Pierre Laperdrix, Nataliia Bielova, Benoit Baudry, and Gildas Avoine. 2020. Browser Fingerprinting: A survey. *ACM Transactions on the Web* 14, 2 (2020). https://doi.org/10.1145/3386040

[91] Annabel Latham and Sean Goltz. 2019. A Survey of the General Public's Views on the Ethics of Using AI in Education. In *Proceedings of the 20th International Conference on Artificial Intelligence in Education (AIED '19)*, Vol. 11625. Springer, 194–206. https://doi.org/10.1007/978-3-030-23204-7_17

[92] Nicholas LePan. 2020 (accessed August 24, 2020). Visualizing the Length of the Fine Print, for 14 Popular Apps. https://www.visualcapitalist.com/terms-of-service-visualizing-the-length-of-internet-agreements/.

[93] Timothy Libert, Lucas Graves, and Rasmus Kleis Nielsen. 2018. *Changes in third-party content on European news websites after GDPR.* Technical Report. Reuters Institute for the Study of Journalism.

[94] Thomas Linden, Rishabh Khandelwal, Hamza Harkous, and Kassem Fawaz. 2020. The Privacy Policy Landscape After the GDPR. In *Proceedings on Privacy Enhancing Technologies Symposium (PETS '20)*, Vol. 2020. De Gruyter, 47–64. https://doi.org/10.2478/popets-2020-0004

[95] Abhishek Mahindrakar and Karuna Pande Joshi. 2020. Automating GDPR Compliance using Policy Integrated Blockchain. In *Proceedings of the 2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity '20), IEEE Intl Conference on High Performance and Smart Computing (HPSC '20), and IEEE Intl Conference on Intelligent Data and Security (IDS '20)*. IEEE, 86–93. https://doi.org/10.1109/BigDataSecurity-HPSC-IDS49724.2020.00026

[96] Célestin Matte, Nataliia Bielova, and Cristiana Santos. 2020. Do Cookie Banners Respect my Choice? : Measuring Legal Compliance of Banners from IAB Europe's Transparency and Consent Framework. In *Proceedings of the 2020 IEEE Symposium on Security and Privacy (SP '20)*. IEEE, 791–809. https://doi.org/10.1109/SP40000.2020.00076

[97] Roman Matzutt, Jens Hiller, Martin Henze, Jan Henrik Ziegeldorf, Dirk Müllmann, Oliver Hohlfeld, and Klaus Wehrle. 2018. A Quantitative Analysis of the Impact of Arbitrary Blockchain Content on Bitcoin. In *Proceedings of the 22th International Conference on Financial Cryptography and Data Security (FC '18)*, Vol. 10957. Springer, 420–438. https://doi.org/10.1007/978-3-662-58387-6_23

[98] Roman Matzutt, Benedikt Kalde, Jan Pennekamp, Drichel Arthur, Martin Henze, Thomas Bergs, and Klaus Wehrle. 2020. How to Securely Prune Bitcoin's Blockchain. In *Proceedings of the 19th IFIP Networking 2020 Conference (NETWORKING '20)*. IEEE.

[99] Aleecia M. McDonald and Lorrie Faith Cranor. 2008. The Cost of Reading Privacy Policies. *I/S: A Journal of Law and Policy for the Information Society* 4, 3 (2008), 543–568.

[100] Maryam Mehrnezhad. 2020. A Cross-Platform Evaluation of Privacy Notices and Tracking Practices. In *Proceedings of the 2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW '20)*. IEEE, 97–106. https://doi.org/10.1109/EuroSPW51379.2020.00023

[101] Ricardo Mendes and João P. Vilela. 2017. Privacy-Preserving Data Mining: Methods, Metrics, and Applications. *IEEE Access* 5 (2017), 10562–10582. https://doi.org/10.1145/3289600.3291384

[102] Anthony D. Miyazaki. 2008. Online Privacy and the Disclosure of Cookie Use: Effects on Consumer Trust and Anticipated Patronage. *Journal of Public Policy & Marketing* 27, 1 (2008), 19–33. https://doi.org/10.1509/jppm.27.1.19

[103] Itishree Mohallick, Katrien De Moor, Özlem Özgöbek, and Jon Atle Gulla. 2018. Towards New Privacy Regulations in Europe: Users' Privacy Perception in Recommender Systems. In *Proceedings of the 11th International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage (SpaCCS '18)*, Vol. 11342. Springer, 319–330. https://doi.org/10.1007/978-3-030-05345-1_27

[104] Jayashree Mohan, Melissa Wasserman, and Vijay Chidambaram. 2019. Analyzing GDPR Compliance Through the Lens of Privacy Policy. In *Proceedings of the VLDB Workshop on Data Management and Analytics for Medicine and Healthcare (DMAH/Poly '19)*, Vol. 11721. Springer, 82–95. https://doi.org/10.1007/978-3-030-33752-0_6

[105] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. 2020. Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM. https://doi.org/10.1145/3313831.3376321

[106] NOYB – European Center for Digital Rights. 2020. My Privacy is None of Your Business. https://noyb.eu/en.

[107] Md Mehedi Hassan Onik, Chul-Soo Kim, Nam-Yong Lee, and Jinhong Yang. 2019. Privacy-aware blockchain for personal data sharing and tracking. *Open Computer Science* 9 (2019), 80–91. https://doi.org/10.1515/comp-2019-0005

[108] Vishwas T. Patil and R. K. Shyamasundar. 2018. Efficacy of GDPR's Right-to-be-Forgotten on Facebook. In *Proceedings of the 14th International Conference on Information Systems Security (ICISS '18)*, Vol. 11281. Springer, 364–385. https://doi.org/10.1007/978-3-030-05171-6_19

[109] Eugenia Politou, Efthimios Alepis, and Constantinos Patsakis. 2018. Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. *Journal of Cybersecurity* 4, 1 (2018), 1–20. https://doi.org/10.1093/cybsec/tyy001

[110] Enric Pujol, Oliver Hohlfeld, and Anja Feldmann. 2015. Annoyed Users: Ads and Ad-Block Usage in the Wild. In *Proceedings of the 2015 ACM Internet Measurement Conference (IMC '15)*. ACM, 93–106. https://doi.org/10.1145/2815675.2815705

[111] Quantcast. 2018 (accessed April 12, 2020). Simplify User Consent. https://www.quantcast.com/gdpr/consent-management-solution/.

[112] Ashwini Rao and Juergen Pfeffer. 2019. Data Siphoning Across Borders: The Role of Internet Tracking. In *Proceedings of the 2019 1st IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA '19)*. IEEE, 168–176. https://doi.org/10.1109/TPS-ISA48467.2019.00028

[113] Ashwini Rao, Florian Schaub, Norman Sadeh, Alessandro Acquisti, and Ruogu Kang. 2016. Expecting the Unexpected: Understanding Mismatched Privacy Expectations Online. In *Proceedings of the 12th Symposium on Usable Privacy and Security (SOUPS '16)*. USENIX Association, 77–96.

[114] Simone Raponi and Roberto Di Pietro. 2020. A Longitudinal Study on Web-Sites Password Management (in) Security: Evidence and Remedies. *IEEE Access* 8 (2020), 52075–52090. https://doi.org/10.1109/ACCESS.2020.2981207

[115] Simon Rebiger. 2016 (accessed April 12, 2020). EU-Parlament beschließt Datenschutzgrundverordnung. https://netzpolitik.org/2016/eu-parlament-beschliesst-datenschutzgrundverordnung/.

[116] Charles Riley. 2019 (accessed April 18, 2020). UK proposes another huge data fine. This time, Marriott is the target. https://edition.cnn.com/2019/07/09/tech/marriott-data-breach-fine/index.html.

[117] Mathieu Rosemain. 2019 (accessed April 12, 2020). France fines Google $57 million for European privacy rule breach. https://www.reuters.com/article/us-google-privacy-france/idUSKCN1PF208.

[118] Takahito Sakamoto and Masahiro Matsunaga. 2019. After GDPR, Still Tracking or Not? Understanding Opt-Out States for Online Behavioral Advertising. In *Proceedings of the 2019 IEEE Symposium on Security and Privacy Workshops (SPW '19)*. IEEE, 92–99. https://doi.org/10.1109/SPW.2019.00027

[119] Iskander Sanchez-Rola, Matteo Dell'Amico, Platon Kotzias, Davide Balzarotti, Leyla Bilge, Pierre-Antoine Vervier, and Igor Santos. 2019. Can I Opt Out Yet?: GDPR and the Global Illusion of Cookie Control. In *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security (ASIACCS '19)*. ACM, 340–351. https://doi.org/10.1145/3321705.3329806

[120] Adam Satariano. 2019 (accessed April 12, 2020). Google Is Fined $57 Million Under Europe's Data Privacy Law. https://www.nytimes.com/2019/01/21/technology/google-europe-gdpr-fine.html.

[121] Adam Satariano. 2019 (accessed April 18, 2020). After a Data Breach, British Airways Faces a Record Fine. https://www.nytimes.com/2019/07/08/business/british-airways-data-breach-fine.html.

[122] Paul Sawers. 2019 (accessed April 18, 2020). 5 data privacy startups cashing in on GDPR. https://venturebeat.com/2019/07/23/5-data-privacy-startups-cashing-in-on-gdpr/.

[123] Lennard Schmidt, Rico Bornschein, and Erik Maier. 2020. The effect of privacy choice in cookie notices on consumers' perceived fairness of frequent price changes. *Psychology & Marketing* 37, 9 (2020), 1263–1276. https://doi.org/10.1002/mar.21356

[124] Seth Schoen. 2009 (accessed April 12, 2020). New Cookie Technologies: Harder to See and Remove, Widely Used to Track You. https://www.eff.org/deeplinks/2009/09/new-cookie-technologies-harder-see-and-remove-wide.

[125] Mark Scott, Laurens Cerulus, and Kayali Laura. 2018 (accessed April 12, 2020). Six months in, Europe's privacy revolution favors Google, Facebook. https://www.politico.eu/article/gdpr-facebook-google-privacy.

[126] Felix Sebastian. 2019 (accessed June 21, 2020). GDPR in the US: Requirements for US Companies. https://termly.io/resources/articles/gdpr-in-the-us/.

[127] Andrew Shindler. 2019 (accessed June 28, 2020). Enforcement of the GDPR in North America — The Experience So Far. https://www.jdsupra.com/legalnews/enforcement-of-the-gdpr-in-north-13166/.

[128] Than Htut Soe, Oda Elise Nordberg, Frode Guribye, and Marija Slavkovik. 2020. Circumvention by design-dark patterns in cookie consent for online news outlets. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI '20)*. ACM. https://doi.org/10.1145/3419249.3420132

[129] Jannick Sørensen and Sokol Kosta. 2019. Before and After GDPR: The Changes in Third Party Presence at Public and Private European Websites. In *Proceedings of the 28th World Wide Web Conference on World Wide Web (WWW '19)*. IW3C2, 1590–1600. https://doi.org/10.1145/3308558.3313524

[130] Jannick Kirk Sørensen and Hilde Van den Bulck. 2020. Public service media online, advertising and the third-party user data business: A trade versus trust dilemma? *Convergence* 26, 2 (2020), 421–447. https://doi.org/10.1177/1354856518790203

[131] Jannick Kirk Sørensen, Hilde Van den Bulck, and Sokol Kosta. 2020. Stop Spreading The Data: PSM, Trust, and Third-Party Services. *Journal of Information Policy* 10 (2020), 474–513. https://doi.org/10.5325/jinfopoli.10.2020.0474

[132] Martin Stabauer. 2019. The Effects of Privacy Awareness and Content Sensitivity on User Engagement. In *Proceedings of the 6th International Conference on HCI in Business, Government and Organizations. Information Systems and Analytics (HCIBGO '19)*, Vol. 11589. Springer, 242–255. https://doi.org/10.1007/978-3-030-22338-0_20

[133] StatCounter. 2020 (accessed August 24, 2020). Browser Market Share Worldwide. https://gs.statcounter.com/browser-market-share.

[134] Statista Research Department. 2015 (accessed May 16, 2020). Europe: daily time spent online via mobile 2015, by age group. https://www.statista.com/statistics/433844/daily-time-spent-online-mobile-age-europe/.

[135] Katarzyna Szymielewicz and Bill Budington. 2018 (accessed April 12, 2020). The GDPR and Browser Fingerprinting: How It Changes the Game for the Sneakiest Web Trackers. https://www.eff.org/de/deeplinks/2018/06/gdpr-and-browser-fingerprinting-how-it-changes-game-sneakiest-web-trackers.

[136] Sven Taylor. 2019 (accessed August 24, 2020). Firefox Privacy – The Complete How-To Guide. https://restoreprivacy.com/firefox-privacy/.

[137] Omer Tene, Katrine Evans, Bruno Gencarelli, Gabe Maldoff, and Gabriela Zanfir-Fortuna. 2019. GDPR at Year One: Enter the Designers and Engineers. *IEEE Security & Privacy* 17, 6 (2019), 7–9. https://doi.org/10.1109/MSEC.2019.2938196

[138] Welderufael B. Tesfay, Peter Hofmann, Toru Nakamura, Shinsaku Kiyomoto, and Jetzabel Serna. 2018. I Read but Don't Agree: Privacy Policy Benchmarking using Machine Learning and the EU GDPR. In *Proceedings of the 27th International Conference Companion on World Wide Web (WWW '18 Companion)*. IW3C2, 163–166. https://doi.org/10.1145/3184558.3186969

[139] Welderufael B Tesfay, Peter Hofmann, Toru Nakamura, Shinsaku Kiyomoto, and Jetzabel Serna. 2018. PrivacyGuide: Towards an Implementation of the EU GDPR on Internet Privacy Policy Evaluation. In *Proceedings of the 4th ACM International Workshop on Security and Privacy Analytics (IWSPA '18)*. ACM, 15–21. https://doi.org/10.1145/3180445.3180447

[140] The Internet Archive. 1996 (accessed August 20, 2020). About the Internet Archive. https://archive.org/about/.

[141] The Washington Post. 2020 (accessed August 6, 2020). The Washington Post. https://www.washingtonpost.com/gdpr-consent/.

[142] Catherine Thorbecke. 2019 (accessed April 18, 2020). Facebook agrees to pay UK fine over Cambridge Analytica scandal while admitting no liability. https://abcnews.go.com/Business/facebook-agrees-pay-uk-fine-cambridge-analytica-scandal/story?id=66635145.

[143] John P. Tomaszewski. 2017 (accessed March 18, 2021). "Opening Clauses" and the GDPR — It Might Not Be As Easy As We Thought. https://www.globalprivacywatch.com/2017/07/opening-clauses-and-the-gdpr-it-might-not-be-as-easy-as-we-thought/.

[144] Moritz Tremmel. 2019 (accessed April 12, 2020). Browser-Fingerprinting gestern und heute. https://www.golem.de/news/datenschutz-browser-fingerprinting-gestern-und-heute-1906-142013.html.
[145] Martino Trevisan, Stefano Traverso, Eleonora Bassi, and Marco Mellia. 2019. 4 Years of EU Cookie Law: Results and Lessons Learned. *Proceedings on Privacy Enhancing Technologies Symposium (PETS '19)* 2019, 2 (2019), 126–145. https://doi.org/10.2478/popets-2019-0023
[146] Tobias Urban, Martin Degeling, Thorsten Holz, and Norbert Pohlmann. 2019. "Your Hashed IP Address: Ubuntu.": Perspectives on Transparency Tools for Online Advertising. In *Proceedings of the 35th Annual Computer Security Applications Conference (ACSAC '19)*. ACM, 702–717. https://doi.org/10.1145/3359789.3359798
[147] 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 ESORICS 2019 International Workshops on Data Privacy Management and Cryptocurrencies and Blockchain Technology (DPM/CBT '19)*, Vol. 11737. Springer, 61–79. https://doi.org/10.1007/978-3-030-31500-9_5
[148] Tobias Urban, Dennis Tatang, Martin Degeling, Thorsten Holz, and Norbert Pohlmann. 2020. Measuring the Impact of the GDPR on Data Sharing in Ad Networks. In *Proceedings of the 2020 ACM Asia Conference on Computer and Communications Security (ASIACCS '20)*. ACM. https://doi.org/10.1145/3320269.3372194
[149] 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 SIGSAC Conference on Computer and Communications Security (CCS '19)*. ACM, 973–990. https://doi.org/10.1145/3319535.3354212
[150] Pelayo Vallina, Álvaro Feal, Julien Gamba, Narseo Vallina-Rodriguez, and Antonio Fernández Anta. 2019. Tales from the Porn: A Comprehensive Privacy Analysis of the Web Porn Ecosystem. In *Proceedings of the 2019 ACM Internet Measurement Conference (IMC '19)*. ACM, 245–258. https://doi.org/10.1145/3355369.3355583
[151] G. Gultekin Varkonyi, Attila Kertész, and Sz. Varadi. 2019. Privacy-awareness of Users in our Cloudy Smart World. In *Proceedings of the 2019 4th International Conference on Fog and Mobile Edge Computing (FMEC '19)*. IEEE, 189–196. https://doi.org/10.1109/FMEC.2019.8795310
[152] Verbraucherzentrale Bundesverband e.V. und die Verbraucherzentralen. 2019 (accessed April 12, 2020). Soziale Medien und die DSGVO: Recht auf Auskunft und Datenübertragbarkeit. https://www.marktwaechter.de/digitale-welt/marktbeobachtung/soziale-medien-und-die-dsgvo-recht-auf-auskunft-und-datenuebertragbarkeit.
[153] Natalija Vlajic, Marmara El Masri, Gianluigi M Riva, Marguerite Barry, and Derek Doran. 2018. Online Tracking of Kids and Teens by Means of Invisible Images: COPPA vs. GDPR. In *Proceedings of the 2nd International Workshop on Multimedia Privacy and Security (MPS '18)*. ACM, 96–103. https://doi.org/10.1145/3267357.3267370
[154] Ben Weinshel, Miranda Wei, Mainack Mondal, Euirim Choi, Shawn Shan, Claire Dolin, Michelle L Mazurek, and Blase Ur. 2019. Oh, the Places You've Been! User Reactions to Longitudinal Transparency About Third-Party Web Tracking and Inferencing. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (CCS '19)*. ACM, 149–166. https://doi.org/10.1145/3319535.3363200
[155] Bjørn Wessel-Tolvig. 2020 (accessed May 16, 2020). What is legitimate interest under the GDPR? https://cookieinformation.com/resources/blog/what-is-legitimate-interest-under-the-gdpr.
[156] Klaus Wiedemann. 2020. The ECJ's Decision in "Planet49" (Case C-673/17): A Cookie Monster or Much Ado About Nothing? *IIC-International Review of Intellectual Property and Competition Law* 51 (2020), 543–553. https://doi.org/10.1007/s40319-020-00927-w
[157] Shomir Wilson, Florian Schaub, Frederick Liu, Kanthashree Mysore Sathyendra, Daniel Smullen, Sebastian Zimmeck, Rohan Ramanath, Peter Story, Fei Liu, Norman Sadeh, et al. 2018. Analyzing privacy policies at scale: From crowdsourcing to automated annotations. *ACM Transactions on the Web* 13, 1 (2018). https://doi.org/10.1145/3230665
[158] Christian Wirth and Michael Kolain. 2018. Privacy by BlockChain Design: A Blockchain-enabled GDPR-compliant Approach for Handling Personal Data. In *Proceedings of 1st ERCIM Blockchain Workshop 2018*. EUSSET. https://doi.org/10.18420/blockchain2018_03
[159] Andy Wolber. 2019 (accessed August 24, 2020). The 5 Best Private Web Browsers of 2019. https://www.lifewire.com/best-private-web-browsers-4177138.
[160] Sebastian Zimmeck and Steven M. Bellovin. 2014. Privee: An Architecture for Automatically Analyzing Web Privacy Policies. In *Proceedings of the 23rd USENIX Security Symposium (SEC '14)*. USENIX Association, 1–16.

A GLOSSARY

In this section, we give brief definitions of terms and concepts relevant to understanding the foundation of our manuscript. Furthermore, in other contexts, certain terms (e.g., fingerprinting) can have a different meaning.

CMPs / Consent Management Platforms In the most general sense, Consent Management Platforms constitute a type of software product. More specifically, CMPs are libraries implementing functionality related to personal data processing. In our scope, their main use is to achieve conformity with legal requirements, most notably, informing end-users that their personal data is processed and providing them with the opportunity to object. Furthermore, we assume that a CMP either directly implements or at least supplies supporting functionality to implement a cookie banner that satisfies said use.

Cookie For our purposes, we refer to cookies as key-value pairs that are strictly associated with a specific website, stored by an end-user's web browser for a given time. The web browser includes the cookies in each (subsequent) request to the website. Thus, in general, cookies allow the web server to persist information on the client's device, i.e., to remember the state, but the central use w.r.t. our work is the identification of a data subject.

Cookie Syncing The practice of associating multiple distinct cookies with one another is referred to as cookie syncing. If at least one of the cookies is used to identify an individual, all cookies can thus be linked to this individual.

Cookie Banner Cookie banners are website (HTML) elements, specifically designed to comply with the requirements of transparency and explicit consent as defined by the ePrivacy Directive and GDPR in relation to personal data processing. We make no further assumptions regarding their form or function other than that their purpose generally consists of informing end-users of the usage of and potentially providing control to avoid trackers.

Cookie Consent Notice See **Cookie Banner**.

Directive A directive within the EU sets specific goals that are to be achieved through national laws, formulated, passed, and enacted by the member states it concerns, respectively.

Fingerprint(ing) We define fingerprinting as the practice of processing any combination of application, system, or device parameters with the explicit purpose of identifying an individual. Therefore, processing a vast quantity of browser-specific settings does not qualify as fingerprinting, if the (in)ability to differentiate between multiple distinct individuals is irrelevant in evaluating the function of the processing system. We make no clear differentiation between device and browser fingerprinting (since we primarily focus on the web), but we want to point out that such a distinction is sometimes made and might be meaningful, depending on the circumstance.

Privacy Policy If a (web) service processes personal information of its users, it is legally required to disclose certain aspects concerning this processing to these users. We regard every text that a web service hosts specifically to satisfy this requirement as (part of the) privacy policy.

Regulation An EU regulation puts the restrictions, obligations, and rights specified within it into EU-wide law. As such, individuals can claim these rights directly, regardless of national laws, unless explicitly specified otherwise in the respective regulation.

Scopus An online search engine developed by *Elsevier Inc.* that indexes academic works of various scientific subjects.

Tracker We use the term tracker to refer to any means of potential identification of an individual. Prominent examples include cookies and fingerprints, but the term is not limited to those concepts. Furthermore, we do not assume the purpose of a tracker. Instead, it is only defined through its potential to identify users. Consequently, we also include purely functional applications of means to identify users, such as authentication (which can even be performed by third parties). However, this definition more accurately represents the privacy risk associated with these means, since the individual being tracked has no control over how data obtained through trackers (even purely functional ones) is utilized.

Tracking The term tracking refers to the application of trackers.

Source	Measurement	Year-Month	Value	Dataset Size	Inclusion Criteria
Cookies					
		Average Number			
[102]	cookies	2000-01	1.57	406	popular
[102]	cookies	2007-02	3.84	406	popular
[42]	cookies	2016-07	17.70	1000000	popular
[145]	cookies	2017-04	24.71	35862	popular EU news
[130]	cookies	2017-08	42.95	64	news
[74]	cookies	2018-02	37.00	500	popular UK
[129]	cookies	2018-02	11.31	3128	miscellaneous
[131]	cookies	2018-02	28.00	342	popular news
[81]	cookies	2018-05	16.40	28227	popular EU
[81]	cookies	2018-06	12.45	28227	popular EU
[119]	cookies	2018-07	92.30	2000	popular EU
[129]	cookies	2018-09	9.67	3128	miscellaneous
[81]	cookies	2018-12	14.59	28227	popular EU
[74]	cookies	2019-02	32.00	500	popular UK
[78]	cookies	2019-02	22.30	1391	popular EU
[131]	cookies	2019-10	23.50	342	popular news
		Legal Compliance			
[145]	cookies problematic	2017-04	49.00 %	35862	popular EU news
[25]	cookies problematic	2018-01	46.00 %	100000	popular
Cookie Banners					
		Adoption			
[145]	banners	2017-04	43.66 %	300	popular EU
[31]	banners	2018-01	46.10 %	9044	popular EU
[74]	banners	2018-05	77.00 %	100	popular UK
[31]	banners	2018-05	62.10 %	9044	popular EU
[119]	banners	2018-07	60.00 %	2000	popular
[16]	banners	2018-11	64.00 %	10000	popular EU
[150]	banners	2018-12	4.00 %	6346	porn
[100]	banners	2020-04	91.00 %	116	popular
		Legal Compliance			
[96]	banners problematic	2019-08	54.00 %	560	popular EU
[105]	banners problematic	2019-09	87.80 %	680	popular UK
[128]	banners problematic	2020-04	5.00 %	300	news

Table 2. We extracted various data points from the respective related work. *Cookies* denotes the average number of cookies per site within the respective dataset. *Banners* denotes the adoption rate of cookie banners within the respective dataset. Measurements denoted as *problematic* indicate the prevalence of potential violations of EU privacy legislation within the respective dataset.

Table continues on the next page.

B RAW NUMBERS FROM THE CONSIDERED MEASUREMENT STUDIES

In Table 2, we further give some detailed insights into the measurement studies conducted by related work that we initially presented in our survey summary in Table 1 and discussed as well as analyzed in our manuscript.

Source	Measurement	Year-Month	Value	Dataset Size	Inclusion Criteria
Privacy Policies					
		Length			
[80]	policy length	2003-08	2806.30	50	popular US
[99]	policy length	2005-10	2565.66	75	popular
[94]	policy length	2016-01	1936.15	22114	popular
[31]	policy length	2016-03	2145.00	7812	popular EU
[31]	policy length	2018-03	3044.00	7812	popular EU
[119]	policy length	2018-07	2261.00	2000	popular
[94]	policy length	2019-05	2621.38	22114	popular
[79]	policy length	2020-07	1916.33	284	popular SA
		Complexity			
[80]	policy complexity	2003-08	14.21	50	popular US
[119]	policy complexity	2018-07	10.50	2000	popular
[79]	policy complexity	2020-07	12.40	284	popular South Asian
Fingerprinting Prevalence					
		All			
[42]	fingerprinting	2016-07	1.60 %	1000000	popular
[3]	fingerprinting	2018-08	68.8 %	10000	popular
[150]	fingerprinting	2018-12	5.00 %	6346	porn
[100]	fingerprinting	2020-04	63.00 %	3695	popular
		Canvas-Based			
[54]	fingerprinting canvas	2015-06	0.85 %	10000	popular
[42]	fingerprinting canvas	2016-07	3.19 %	1000000	popular
		Flash-Based			
[2]	fingerprinting flash	2013-11	0.95 %	10000	popular
[54]	fingerprinting flash	2015-06	0.63 %	10000	popular
		Font-Based			
[2]	fingerprinting font	2013-11	0.04 %	10000	popular
[54]	fingerprinting font	2015-06	0.22 %	10000	popular
[42]	fingerprinting font	2016-07	1.40 %	1000000	popular

Table 2. We extracted various data points from the respective related work. *Policy length* denotes the average number of words within the respective dataset. *Policy complexity* denotes the average Flesch-Kincaid score within the respective dataset. *Fingerprinting* denotes the prevalence of the denoted fingerprinting technique within the respective dataset. Measurements denoted as *problematic* indicate the prevalence of potential violations of EU privacy legislation within the respective dataset.