ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Sivan-Sevilla, Ido; Parham, Patrick; McGuigan, Lee

Article

"Cookie-less" identification for/against privacy?

Internet Policy Review

Provided in Cooperation with:
Alexander von Humboldt Institute for Internet and Society (HIIG), Berlin

Suggested Citation: Sivan-Sevilla, Ido; Parham, Patrick; McGuigan, Lee (2025) : "Cookie-less" identification for/against privacy?, Internet Policy Review, ISSN 2197-6775, Alexander von Humboldt Institute for Internet and Society, Berlin, Vol. 14, Iss. 3, pp. 1-27, https://doi.org/10.14763/2025.3.2025

This Version is available at:
https://hdl.handle.net/10419/324163

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.


https://creativecommons.org/licenses/by/3.0/de/deed.en

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

INTERNET POLICY REVIEW

Volume 14 **Issue** 3

RESEARCH ARTICLE

OPEN ACCESS

PEER REVIEWED

‘Cookie-less’ identification for/against privacy?

Ido Sivan-Sevilla *University of Maryland*
Patrick Parham *University of Maryland*
Lee McGuigan *University of North Carolina at Chapel Hill*

DOI: https://doi.org/10.14763/2025.3.2025

Published: 6 August 2025
Received: 19 September 2024 **Accepted:** 4 February 2025

Funding: The authors did not receive any funding for this research.
Competing Interests: The author has declared that no competing interests exist that have influenced the text.
Licence: This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License (Germany) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. https://creativecommons.org/licenses/by/3.0/de/deed.en
Copyright remains with the author(s).

Citation: Sivan-Sevilla, I., Parham, P., & McGuigan, L. (2025). ‘Cookie-less’ identification for/against privacy? *Internet Policy Review*, *14*(3). https://doi.org/10.14763/2025.3.2025

Keywords: Cookie-less advertising, Consumer privacy, Online advertisement, Commercial surveillance

Abstract: The advertising industry’s anticipated shift away from third-party cookies led to the proliferation and normalisation of first-party identification architectures online. Marketed as ‘privacy-friendly,’ the new technologies promise to deliver the efficiencies that advertisers have become accustomed to, while addressing privacy concerns from third-party cookies. Such tension calls for a better understanding of the privacy implications from first-party online identification architectures. We evaluate first-party user identification mechanisms by (1) surveying the literature to create a typology that synthesises existing privacy concerns in third-party cookie-based identification, and (2) applying our typology to evaluate the privacy of prime examples in what we frame as three distinct types of first-party ID architectures — Universal IDs, Onboarding ID, and Walled Garden ID. We analyse technical documentation and code repositories from each architecture type and show how first-party ID solutions still enable cross-site tracking over longer periods of time and encourage sensitive user targeting. First-party ID solutions do create mechanisms to ease opting out from tracking, but the implementation of those mechanisms is questionable. Our findings demonstrate how the advertising industry is trying to maintain its existing structure and replicate the tracking functionalities on which it has grown reliant.

1 - Introduction

Continuous identity-building of online consumers has been the crown jewel of the digital advertising industry (Alaimo, 2022; Gandy, 2021; Mellet & Beauvisage, 2020). By using individual data points that go beyond identification details, advertising actors have been intensively crafting user identities for profiling purposes, without user awareness (Chant, 2021; EU Commission, 2023), violating consumers' privacy right to freely and continuously reconstruct the self in the face of ever-changing contexts (Agre & Rotenberg, 1997). Taking control over the construction of consumer selves (Benoist, 2008; Elmer, 2003; Zwick & Dholakia, 2004), companies that facilitated personalised advertising made themselves integral to the monetisation of the internet (Crain, 2021; Kant, 2021; Turow, 2011). Identity services became infrastructural to online advertising markets that, in the US alone, generated USD 225 billion in revenue in 2023 (IAB, 2024).

The pervasive identity-building of consumers has been enabled by *online identification architectures*. These architectures allow continuous attribution of collected data to specific users and feed the representation of individuals as correlated data subjects, based on past behaviour, often without the knowledge of the data subjects themselves (Hildebrandt, 2006). Online identification emerged shortly after the birth of the Internet. Unique user identifiers in the form of third-party (3p) cookies became the most popular anchors for collected user data (Turow, 2011). Originally designed to enable continuities in browsing experiences (namely, letting products placed in online shopping carts remain there while a user navigated to other pages, to continue shopping, perhaps), third-party cookies became the go-to instrument for identifying online sessions and tracking consumers' behaviour for identity-building practices (Jones, 2020). Advertising actors have been using third-party cookie identifiers by dropping them on browsers for persistent consumer identification across the web (Sivan-Sevilla & Poudel 2025), the collection of browsing history, and to conduct cookie syncing practices for linking consumer identities across devices and between different advertising actors (Solomos et al., 2020). They are called *third-party* cookies because the companies deploying them do not have a direct (or *first-party*) relationship with the user—the presumption being that first-parties, like the websites or merchants a user consciously engages with, are legitimately authorised to collect user data, whereas third-parties have less authorisation.

Recently under pressure to reform online privacy problems, Apple and Mozilla have restricted the use of 3p cookies on their browsers, while Google has repeatedly made, and then withdrawn, that same pledge (Morgan & Mazalon, 2024). The

looming inability to use third-party cookies for user identification created a vacuum for advertising actors. Without attributing collected data to consumers, advertising revenue was perceived to be at risk (Brodherson et al., 2021). Our study focuses on one highly-publicised method for filling this vacuum: first-party identification architectures, which are marketed by the Interactive Advertising Bureau (IAB), the leading advertising industry trade association, as 'privacy first solutions' (Eng, 2024; IAB Canada, 2021). Intriguingly, what were previously marginal consumer identification mechanisms have become increasingly popular and normalised as 'post-cookies' & 'privacy-first' user identification based on first-party user data (Meltzer, 2020; Tabisz, 2023). The advertising industry calls these "universal" or "cookie-less" identifiers; we suggest that a more fitting name is "first-party identification architectures," and ask, what are the privacy implications of first-party identification architectures?

'First-party identification architectures' provide a means for the digital advertising industry to legitimise and maintain certain abilities to identify consumers. First-party identification architectures are distinguished from prior third-party systems based on the source of data and identification mechanisms; in the former, companies within digital advertising markets (e.g., advertisers, adtech intermediaries, and publishers) supply their own proprietary data, including emails, names, and phone numbers, to find consumers across ad-supported media venues, rather than relying on the passive surveillance executed via third-party cookies. This name is somewhat imprecise, in that these architectures still rely on the sharing of first-party data between different digital advertising ecosystem participants; the industry's self-serving use of 'first-party' glosses over shifting classifications of what is first-party versus third-party data in these contexts (McGuigan et al., 2023). Nevertheless, we suggest that this term effectively captures the nature of the project—of using the legitimising connotation of first-party relationships to authorise continued tracking and identification. We argue that three types of first-party identification architectures have emerged to help facilitate the identification of consumers from different positions in the advertising supply chain.

We recognise three distinct first-party identification architectures that have emerged in the online advertising supply chain: (1) Publishers[1] are deploying 'Universal ID' solutions to persistently identify logged-in consumers; (2) Advertising technology partners provide 'Onboarding Identification' mechanisms to link first-party advertiser data with user identity across advertising platforms; and (3)

1. In digital advertising terminology, a publisher is any entity that exhibits content and sells advertising opportunities on its website(s), whether a large news organisation or an independent blogger.

'Walled-Garden Identification' enables dominant actors that are active on both the buyer and seller sides of online advertising, such as Google and Meta, to link their first-party user data across the different products of the same 'walled-garden' platform.

Our paper systematically compares the capabilities and recommended uses of first-party identification architectures to the known surveillance and privacy concerns surrounding third-party cookies, thereby allowing a rigorous assessment of whether the former meaningfully solve—or essentially reproduce—the problems associated with the latter. We first conduct an interdisciplinary review of literature on the privacy concerns of 3p cookie-based identification and tracking. We followed a snowball approach to collect literature until reaching saturation in terms of spotting new themes of privacy issues. We organised the themes that emerged from our review into a novel five-part typology of 3p cookie-based privacy concerns. We then used that typology to evaluate exemplars of each first-party identification architecture: The Trade Desk's Universal ID 2.0 solution, LiveRamp's RampID onboarding identifier, and Google's Customer Match walled-garden identification solution. Our selection of prime examples to analyse was informed by our reading of trade publications covering the digital advertising industry, the frequency of reporting on specific first-party solutions, and the expected adoption rates of those solutions by different publishers and AdTech actors (Asim, 2021, 2022).

All three types of first-party identification architectures are similar in that they rely on the encryption of first-party data to identify users. Still, we observed distinct identification objectives across these solutions: Universal IDs operate across sites within networks of participating publishers; Onboarding identification helps buyers of ads target users in almost any site or platform; and Walled Garden Identification helps buyers target users in specific platform environments. We collected data on each first-party identification case from technical documentation that the developers of the different solutions have made public. For each solution, we analysed the code repositories, related industry publications, media reporting, and relevant documents, in light of our developed evaluation typology of existing privacy concerns with 3p cookie identifiers.

Contrary to industry claims that first-party ID solutions advance consumers' privacy, we find that first-party identification actually intensifies important aspects of online surveillance: the identifiers and the data profiles associated with them can be more persistently attached to individuals; the personal data profiles can be more comprehensive; and the identifiers can allow for sensitive forms of profiling and discrimination to persist and even skirt efforts to disable them. There is nu-

ance to these findings. We do recognise progress regarding the set of actors that can track users and some promising transparency and user agency mechanisms that could potentially increase the accountability of online trackers. Still, the application of our typology suggests some concerning conclusions about the privacy implications of first-party identification architectures.

First-party identification is far from privacy-inducing. These reforms fail to address the structural conditions, incentives, and power relations that remain in contradiction with many legal provisions (Veale & Borgesius, 2022) and any meaningful definition of privacy (Agre & Rotenberg, 1997; Barocas & Nissenbaum, 2014; Wachter, 2019). Consumer tracking, profiling, and targeting are deeply entrenched as norms and an underlying market infrastructure in these architectures.

The next section details our research methodology and its limitations. We then summarise existing literature on the various privacy implications of third-party cookie identifiers, establishing a typology to later assess first-party identification architectures. Section 4 applies our criteria to compare the privacy implications of first-party identification architectures, analysing a prime example from each first-party identification type. Section 5 discusses the implications of our findings and concludes, detailing future research questions.

2 - Methods

The starting point for our data collection was the advertising industry's claims that ‘first-party’ identification architectures improves users’ privacy, as popular browsers shifted away from third party cookies, and the industry sought new user identification mechanisms (Eng, 2024; IAB Canada, 2021; Meltzer, 2020; Tabisz, 2023). We were fairly skeptical, and decided to first synthesise existing privacy concerns in the online advertising supply chain that are driven from the reliance of advertising on third-party cookies.

Data collection started with a survey of existing literature about tracking via third-party cookies and its privacy consequences. We began with the key words of ‘third-party cookies.’ We used ACM’s digital library and the Google Scholar search engines and collected studies from computer science conferences, law and policy studies, and related fields of marketing and media research. The searches yielded 287 articles, which were then enriched by finding additional studies that cite or are cited by works in the initial sample, expanding in a snowball protocol until reaching saturation. We read and filtered these works and ended up with 67 relevant papers that were manually classified based on the privacy themes that emerged in the

context of third-party cookies. Papers were selected for analysis if they discussed any privacy implication from the use of third-party cookies for tracking and targeting purposes. This first data collection step led to the establishment of our typology, which names five distinct privacy concerns enabled by third-party cookie identifiers, articulates their types and definitions, and lists a sample of relevant studies on each privacy issue (see next section).

The second step of data collection evolved around the collection of information on the suggested first-party identification architectures. Based on the expertise and industry experience of one of our co-authors, we broke down suggested first-party identification architectures to three types: Universal ID solutions, On-boarding identifiers, and Walled Garden identifiers. We then selected three technologies to represent the three different types of identification architectures based on: (1) The Trade Desk UID2.0's positioning as the leading industry wide solution that has been tested by significant number of advertisers and seen continued adoption (Asim, 2022; Barber, 2024; Hercher, 2024); (2) LiveRamp RampID's development from the continued leading onboarding firm in the industry (Hercher, 2019; Shields, 2023); and (3) Google Customer Match's position as the solution from largest Adtech firm and platform by market share and revenue (Yuen, 2024).

We then dived into these three identification architectures and collected data from their github repositories, industry publications, and media reporting. Our main sources include official documentation and industry reports from magazines such as *Digiday*, *AdExchanger*, and *eMarketer*. We also looked into the code repositories and websites of the companies that own the solutions and used the materials provided to assess how each solution works in the broader ad supply chain.

We analysed the collected data on each first-party solution to understand its broader privacy implications in the context of existing privacy concerns with third-party cookies. The list of themes that emerged in step #1 served as a useful guideline for understanding the newly proposed solutions, since they address long-term privacy concerns on what is enabled by third-party cookies, and serve as a point of reference to understand our main puzzle - whether cookie-less identification architectures improve or undermine users' privacy.

Importantly, and as explained in the introduction, we follow the framing of the digital advertising industry when considering what to include as 'first-party' identification mechanisms. Instead of relying on passive surveillance conducted by third-party cookies, the first-party solutions under study supply their own proprietary data to create user identifiers and sync those between actors in the ad supply

chain. First-party, in that sense, is not necessarily first-party to the user, but first-party to the actor that identifies the user in the ad supply chain (publisher, ad network, or walled garden service). The industry uses legitimising connotation of 'first-party' relations to justify continued tracking and identification, and we aim to assess whether 'first-party' relations between the sources of the data and the actor that conducts the identification are indeed harmless from a user privacy perspective.

Our literature review is not an exhaustive list of all third-party cookie-based tracking studies. We followed the snowball approach, and relied on our experience and expertise in studying this field, to come up with what we consider the most important privacy implications of third-party cookie advertising. Additional limitations are related to our ability to decode and navigate the details of each first-party ID specification. Despite our efforts to examine and analyse these first-party identification architectures, the available documentation does not clearly disclose all the relevant details of their functionalities and implementation. We tried to verify our understanding with direct questions to the companies but were unable to retrieve more revealing details. Our findings may not have a long shelf life. Any changes to implementation practices could affect the privacy implications of the different solutions.

3 - Literature review: the privacy implications from third-party cookie based advertising

Table 1 below captures an array of different privacy concerns that emerged from analysing the literature on third-party cookies. The first two concerns are technical - the usage of third-party cookies to identify and track users across Web, mobile, and smart technologies has enabled (1) tracking users across different sites, and (2) tracking users over time, as these cookies get synced and persist for long periods of time. The next two concerns are about user rights. Third-party cookie based tracking (3) makes opting-out very difficult for users, with many trackers often ignoring users' preferences, while providing very little, if any, (4) tracker transparency on who exactly tracks the user over an array of sites and apps. A fifth privacy concern that emerged from the literature is the ability of trackers to use sensitive categories for targeting users. This is a consequence of previous affordances enabled by third-party cookies. The sample column in the table represents the most comprehensive empirical works that studied the privacy concern. The paragraphs below summarise and demonstrate each concern from the reviewed literature.

TABLE 1: User Privacy concerns enabled by third-party cookies

PRIVACY CONCERN	TYPE OF CONCERN	DEFINITION	SAMPLE OF RELEVANT LITERATURE
1 - CROSS-SITE TRACKING	TECHNICAL	THE ABILITY TO TRACK USERS ACROSS MULTIPLE SITES BY VARIOUS ACTORS BASED ON THEIR 3P COOKIE IDENTIFIER.	ENGLEHARDT & NARAYANAN, 2016; KARAJ ET AL., 2019; FOUAD ET AL., 2020.
2 - LONGITUDINAL TRACKING	TECHNICAL	THE ABILITY TO TRACK USERS OVER TIME BASED ON THEIR 3P COOKIE IDENTIFIER.	SAMARASINGHE & MANNAN, 2019; PAPADOPOULOS ET AL., 2019; FOUAD ET AL., 2022
3 - OPT-OUT LIMITATIONS	USER RIGHTS	THE EXTENT TO WHICH USERS CAN OPT-OUT FROM 3P COOKIES TRACKING OR OTHERWISE EXERCISE THEIR PREFERENCES.	MATTE ET AL. 2020; HABIB ET AL. 2022; SANCHEZ-ROLA ET AL., 2019; GRASSL ET AL., 2021; TREVISAN ET AL., 2019
4 - LACK OF TRACKER TRANSPARENCY	USER RIGHTS	USERS' INABILITY TO KNOW WHAT DATA IS COLLECTED ABOUT THEM AND BY WHOM.	DEGELING ET AL. 2019; LIBERT & BINNS, 2019; FOUAD ET AL., 2022
5 - SENSITIVITY OF TARGETING	CONSEQUENTIAL	LINKING 3P COOKIE IDENTIFIERS ACROSS ADVERTISING ACTORS TO TARGET INDIVIDUALS BASED ON CATEGORIES SUCH AS RACE, GENDER, AND RELIGION, THAT HAVE SPECIAL STATUS IN LAW AND/OR ARE PROHIBITED BY CORPORATE POLICIES.	ALI ET AL., 2019; WEI ET AL., 2020; BEAUVISAGE ET AL., 2023

3.1 Cross-site tracking

One of the most commonly discussed limitations on users' ability to keep their identity fragmented online is the use of third-party cookie identifiers to identify and track users *across sites*. Scholars have been measuring the volume of cross-site tracking conducted by various actors, revealing tracking levels based on type of

websites (Sivan-Sevilla & Poudel, 2025), dominance and reachability of certain trackers, and practices of syncing and forwarding 3p cookie IDs between trackers for greater visibility on user behaviour (Fouad et al., 2020; Karaj et al., 2019; Libert & Binns, 2019; Macbeth, 2017; Papadopoulos et al., 2019; Roesner et al., 2012; Yang & Yue, 2020). Lerner et al. (2016) showed how cross-site tracking via third-party cookies on the Web has increased in prevalence and complexity between 1996-2016, with 'more sites…giving more third parties the opportunity to track users' (p. 1007).

Cross-site tracking is mostly carried out by a specific set of dominant trackers, with a small number of third-party trackers observing an increasing portion of users' behaviour on the Web (Englehardt & Narayanan, 2016; Lerner et al., 2016). The reach of trackers is the highest for Google, then Amazon and Facebook, and a long tail of Criteo, Microsoft, Twitter, Adobe and others (Solomos et al., 2020). Specifically, Google's cross-site tracking reach is significantly higher than all other actors (Englehardt & Narayanan, 2016; Fouad et al., 2020; Lerner et al., 2016; Samarasinghe & Mannan, 2019).

Studies have found that trackers frequently share third-party cookie IDs with one another, in what has been framed as 'cookie syncing.' On average, a user experiences one cookie syncing per 68 web page requests (Papadopoulos et al., 2019). Interestingly, some cookie-syncing practices promote a 'universal' ID for users, by setting cookie IDs that were previously set by other domains. In 131 cases of third-party cookie ID sharing, Papadopoulos et al. (2019) spotted the sharing of 'ID summaries'–lists of user IDs that other domains used for a particular user–to enable greater visibility on users. Third-party cookies have enabled user tracking across the majority of visited web pages, by various third-parties, and allowed trackers to share user IDs, through cookie syncing and forwarding practices, creating real-time visibility on user behaviour by an increasing amount of trackers.

3.2 Longitudinal tracking

Third-party cookies enable trackers to learn about users over time, creating a *longitudinal* understanding of user interests and intents. *Longitudinal tracking* is enabled by the respawning of third-party cookie IDs. In contrast to 'regular' third-party cookie IDs, studies showed how users can no longer prevent tracking just by deleting their cookies (Fouad et al., 2022). Respawned cookies replicate the user ID based on features of the user's machine such as IP addresses or user's browser agent. Linking stateless and stateful user identifiers, trackers can create longitudinal user profiles by linking users' activity before and after they clean third-party

cookies in their browsers.

Longitudinal tracking is also enabled by the cookie-syncing practices described in section 3.1 On average, a user gets cookie syncs for seven user IDs over the period of a year (Papadopoulos et al., 2019), ensuring that trackers can maintain visibility on the user over time. Interestingly, some of the most dominant advertising trackers – Rubicon, Yahoo, and others – set cookie IDs that remained valid for more than 20 years, violating EU's cookie law (Samarasinghe & Mannan, 2019).

3.3 Opt-out limitations

Relatedly, limitations are also posed on users' ability to comprehend and decide on the level of online tracking they will be subject to online. Complicated third-party cookie-based tracking mechanisms present challenges for even tech-savvy users in their management of privacy online (Acar et al., 2014). Challenges include the deceptive language of tracking consent notices in consent management platforms (CMPs), which then share user's preferences with third-party trackers (Matte et al., 2020). Evaluation of post-GDPR effects has shown that users can still be tracked across almost 90% of popular sites, and deceiving methods are prominently used to coerce users into providing their consent for third-party cookie-based tracking (Sanchez-Rola et al., 2019). When presented with a privacy policy to make a decision to opt out of data practices, readers form an impression that what they have read aligns with reasonable expectations and not with what is actually in the text (Martin, 2015).

Research on the effectiveness of the top five CMPs has demonstrated that only 11.8% meet the minimum compliance requirements under the GDPR (Nouwens et al., 2020). Further, the third-party cookie consent banners employ "dark patterns" through design choices such as defaults, aesthetic manipulation, and obstruction that do not allow for users to make meaningful decisions (Graßl et al., 2021). Research on the ideal version of cookie banners and consent interfaces found that when consent mechanisms are complex, technical, and ambiguous users are more likely to consent to disclose more information, without understanding why (Habib et al., 2022). Regardless of how consent mechanisms are presented, research has demonstrated that in 49 percent of EU websites, a cookie is placed before a user gives consent to data collection (Trevisan et al., 2019), questioning the ability for users to have meaningful agency to decide on the levels of third-party cookie-based identification & tracking.

3.4 Lack of tracker transparency

The level of *transparency* provided by trackers who engage in third-party cookie based tracking was found to be limited as well. Users cannot follow the dynamics and complexity of third-party cookie-based tracking practices nor can they understand what is collected on them and by whom. They rely on full disclosures by publishers in their privacy policies, but those policies fail to describe observed tracking practices. Libert and Binns (2019), for instance, found a misfit between the way news websites describe tracking practices in their privacy policies and the tracking observed on their sites. Fouad et al. (2022) studied the respawning of cookies and inspected third-party cookie policies in 142 websites to find that none of them describe the respawning behaviour observed in the study.

3.5 User targeting based on sensitive categories

Studies found how advertising agencies enjoy a significant amount of discretion when targeting users and can easily leverage sensitive information to construct audience segments, as detailed below. As advertising agencies continue to embrace big data technologies in audience construction and optimisation, they make considerations based on available demographic data attributes (Beauvisage et al., 2023). Advertisers link third-party cookies to other data sources, allowing buyers of ads to pair their consumer data with third-party cookies, beyond what passive surveillance of browsing behaviour can provide, potentially circumventing restrictions on the targeting of users based on sensitive assembled categories (Sherman, 2021). Data sources for user targeting are barely limited. Previous studies showed how a range of ad actors, online and offline, can contribute to profiling consumers for marketing purposes and identified instances where advertiser-uploaded lists have violated platform policies when targeting users based on categories such as race, religion, politics, sex life, or health (Choi et al., 2020; Wei et al., 2020a).

Our five-criteria typology for the privacy concerns in third-party cookies-based advertising guides our following assessment of first-party identification architectures and their privacy implications.

4 - ‘Privacy-preserving’ first-party identification architectures?

Our analysis of ‘first-party’ identification architectures follows the rhetoric of the advertising industry and considers architectures that rely on data sources that are first party to the identifying actor, which can be a publisher, an ad network, or an

advertising platform. We begin by distinguishing three distinct types of first-party identification architectures (4.1). We then conduct a privacy analysis of each first-party data solution by selecting one suggested application of the ID architecture and evaluating that solution based on our typology of privacy concerns (4.2). We draw comparisons both across solutions and in relation to how the industry is currently using third-party cookie-based user identification (Table 2).

4.1 First-party ID formulation & usage across the ad supply chain

The use of first-party data for the continued identification and targeting of users without the support of third-party cookies is pursued by three different types of identification architectures: 1) Universal IDs, 2) Onboarding identification, and 3) Walled Garden identification.

'Universal IDs' (UIDs) are designed to persistently identify users across a network of participating publishers. They take advantage of the trend toward a "data-wall" business model, wherein publishers restrict access to content to users who create a "free" account and pay with their data (Evens & Van Damme, 2016; Grover & Baik, 2024). The advertising companies, that are US-based companies, are using the American legal concept of Personally Identifiable Information (PII) to describe the type of information from which user tokens will be generated. When a user agrees to login to a publisher's website using PII, such as an email, the UID operator (which could be an advertising or media company) generates an identity token unique to that user. When the user is logged in to their account, they can be recognised on any other website participating in the UID network by the UID operator. Importantly, different publishers and advertisers will receive different tokens for the same user, limiting their own ability of persistent identification.

'Onboarding identification' is an architecture that relies on first-party data sources of the advertising network, and replaces the third-party cookie ID syncing functionality between third-party trackers. This architecture enables buyers of online advertisements to target desired consumers across the various advertising platforms they want to work with (mobile, video, social media sites, and etc.).

'Walled Garden identification' is the means by which dominant actors that are active on both the buyer and seller sides of online advertising allow advertisers to link their first-party user data across the different products owned by the same dominant actor. Google, for instance, identifies users across display, search, and YouTube (Janardhan, 2023), based on first-party data sources it owns on users, enabling advertisers to target users across the different Google products

Figure 1 below, created by the authors, illustrates how different first-party identification architectures come together and serve different purposes across the seller and buyer sides of the online advertising supply chain. The starting point for any first-party identification architecture is the first-party data sources, collected by either sellers (through user logins) or buyers (through past advertising transactions and customer sales data), over time. From the seller's side, participating websites that are part of a 'Universal ID' solution (architecture #1 in the figure) ask their users to log into a personal account, using login details such as an email address. An account log-in initiates the creation of an encrypted identification token by the designated Universal ID operator. These tokens are then used to bid for display ads and target the associated users. From the buyer side, advertisers leverage their first-party data to identify and target customers for ads on other user platforms. Through 'Onboarding identification' solutions (architecture #2 in the figure), buyers can target customers in almost any desired medium, given the breadth of actors who participate in the onboarding identity graph of the ad network. Through 'Walled Garden first-party identification' (architecture #3 in the figure) buyers can target customers in products within the walled garden.

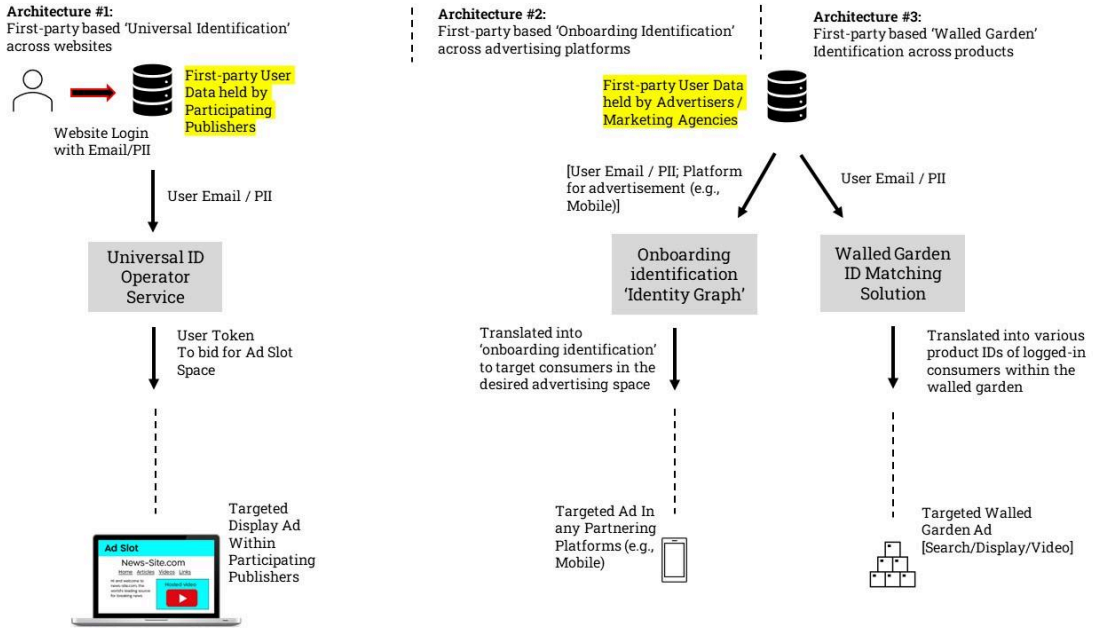

FIGURE 1: First-party identification architectures as enablers of targeting across platforms

We now discuss the three exemplars within these categories - TradeDesk's UID 2.0 as a Universal ID solution, Liveramp's RampID as an 'Onboarding identification' solution, and Google's Customer Match as a 'Walled Garden' solution.

4.1.1 TradeDesk's Unified ID (UID) 2.0 - 'Universal ID'

The solution creates first-party user identification from the seller side (the publishers). It is anchored by the credential(s) users offer when they log in to publishers' websites. Upon initial login, users will be asked to consent to receiving targeted advertisements. If they agree, the publisher connects with the 'UID2.0 Operator,' a service that stores user keys and distributes them to relevant actors. The operator turns the hashed user email or phone number into a 'Raw Unified ID' (Raw UID) with secret salts added.[2]The 'Raw ID' serves as a unique pseudo-anonymous identifier; it is then encrypted to create a 'UID2.0 token' that is further shared with Supply Side Platforms (SSPs) in the bidstream. The usage of tokens instead of raw IDs in the bidstream prevents participating actors on the seller side from viewing the 'raw UID' that could have enabled persistent user tracking. Each entity in the server side will receive a different token for the user from the UID operator. Demand Side Platforms (DSPs) from the buyer side (which bid into auctions on behalf of advertisers) can match encrypted UIDs from SSPs in the bid stream based on decryption keys they get from 'UID Operators.' Overall, the TradeDesk solution creates a PII-based user identifier that is shared among participating actors from the seller (publishers) and buyer (advertisers) side, with the technical possibility for users to opt-out and get more information from UID2.0 operators on the type of user tracking that is taking place.

4.1.2 LiveRamp's RampID - 'Onboarder Identifier'

This solution is distinct from the previous one in that it is triggered by first-party user data from the buyer side. Advertisers and marketing agencies use first-party user identification obtained from their customers and can upload those first-party identifiers to the LiveRamp platform, along with the destination advertising platform in which they want to target users. The first-party user identifier provided by buyers is then matched to a pseudonymous user ID —RampID— within LiveRamp's platform. The RampID is translated into platform-specific identifiers (e.g., mobile device IDs) according to the advertiser's targeting preference (LiveRamp, 2024).

In order to identify users and ultimately generate a RampID, LiveRamp's 'AbiliTech' platform[3]merges multiple sources of data on users, including offline PII and pseudoanonymous identifiers (LiveRamp, 2022c). These offline and online pairings

2. 'Secret salts' is random data that is used as an additional input to hash functions to ensure the original data is not easily decoded.

3. LiveRamp refers to the 'AbiliTech' platform as an 'identity graph.' We refer to 'AbiliTech' as a platform to avoid confusion with how advertisers use identity graphs in other use cases for targeting of users based on sensitive attributes.

are then merged within the AbiliTech platform to represent users (LiveRamp, 2022c). LiveRamp consolidates offline PII about a known individual including: email addresses, postal addresses, phone numbers, and enables resolution where individuals have multiple identifiers in these categories (LiveRamp, 2022c). The LiveRamp algorithm then translates this pairing to a RampID via pseudoanonymous PII from the partners (LiveRamp, 2022c). Offline PII and Online device identifiers can now be matched to represent the user as a RampID, and enable other offline and online data sources about the user to be paired (LiveRamp, 2022c). The matching ID, 'RampID,' is encrypted in the real-time bidding process, when buyers attempt to buy ad space based on the RampID, and can only be decrypted by supply-side platforms (SSPs) and demand-side platforms (DSPs) who are part of the LiveRamp network (LiveRamp, 2022d).

Since RampID is based on first-party data collected or purchased by advertisers, users have very limited visibility or points of intervention in the process. Still, users can choose to opt-out from the RampID solution when they log into websites and share their email address in exchange for content. Their email address can either be connected to a RampID through LiveRamp's API, or disconnected from the RampID solution in case users choose to opt-out (Asim, 2022; LiveRamp, 2022c).

4.1.3 Google's Customer Match / Ads Data Manager - 'Walled Garden Solution'

On October 11, 2023, Google presented its 'Google Ads Data Manager,' a platform that enables the use of first-party identifiers held by advertisers and marketing agencies for the targeting of consumers across Google's walled-garden products. The platform aims to simplify the management of first-party data connections between buyers of online ads and Google's products. According to media reporters, it shows the desire of Google to leverage its massive logged-in user base as a replacement for third-party cookies (Joseph, 2024).

This first-party identification architecture is similar to the previously detailed RampID solution. It is generated from the will of an advertiser / marketing agency to target potential customers based on a set of first-party identifiers held by ad buyers, such as email addresses, phone numbers, names, and home addresses (Janardhan, 2023). Those first-party identifiers are then matched with Google's first-party identifiers of logged-in customers across Google products - Search Engine, Gmail, YouTube, and Google's Display Network (the array of websites where Google exists as a third-party, not necessarily via cookies). The newly introduced Google Ads Data Manager incorporates Google's existing first-party targeting solution, Customer Match, through which the ID matching takes place (Google, n.d.-c, 2024).

In contrast to RampID which can enable advertisers to target users across different advertising platforms, this Google solution is native to Google products, and enables advertisers to find users within Google's (sprawling) walled garden. Advertisers can upload their first-party data by connecting sophisticated first-party data management platforms, such as customer relationship management (CRM) platforms and customer data platforms (CDPs), or through simply uploading a *.csv file of customer data to the Google Ads platform that contains the headers "Email," "Phone," "First Name," "Last Name," "Country," and "Zip" (Google, n.d.-b, n.d.-d, n.d.-f; Janardhan, 2023).

Advertisers also have the option to protect the PII data of their customers from external stakeholders outside the Google network using the SHA256 algorithm for one-way hashing (Google, n.d.-e). Google matches the uploaded PII with corresponding data associated with Google Accounts to create the final Google Match segment that can be used for targeting by Google (Google, n.d.-a).

4.2 Privacy implications of first-party ID architectures

The section below examines the privacy implications of first-party identification architectures, in light of existing privacy concerns from third-party cookies.

4.2.1 Cross-site tracking

The ability to identify users across an increasing number of websites and advertising platforms, and by an array of ad buyers, has been one of the main privacy concerns enabled by third-party cookie identifiers. The different types of first-party identification architectures replicate this technical privacy concern, but limit the set of actors who can do so. Still, the Universal ID operator can persistently identify users across the sites of any publishers participating in the Universal ID solution; users are also persistently identified by buyers of online ads, across different advertising platforms and walled garden products. From the seller's side, cross-site tracking is only enabled on publishers who participate in the Universal ID solution, potentially limiting the amount of actors who get cross-site visibility on users.

UID2.0 enables cross-site tracking only to the UID operator who still gets to view users in the bidstream across their browsing experiences. UID tokens are constantly updated to ensure that unapproved entities with access to the bid stream cannot build profiles based on UID tokens over time. Liveramp's RampID architecture also enables cross-site tracking by matching and sharing Ramp IDs across advertising platforms. Google's Customer Match enables cross-site tracking as well, since advertisers are able to target Google users across third-party publishers that

are part of the Google Display Network. Still, the limits placed on the number of partners able to identify individuals through the introduction of encryption mechanisms could be considered an improvement over the amount of cross-site tracking based on third-party cookie identifiers.

4.2.2 Longitudinal tracking

The ability to learn about user behaviour over time is currently enabled by various third-party cookie-based tracking techniques such as the respawning and syncing of cookie IDs. In first-party identification architectures, this capability is enabled by the introduction of much more persistent identifiers, which are based on PIIs such as user emails or phone numbers. This enables trackers to learn about user behaviour over longer periods of time, based on an identifier that is less likely to change frequently in comparison to third-party cookie IDs, and is often used in other offline data transactions, linking together a rich history of user records.[4]

UID2.0 architecture enables longitudinal tracking for advertising and targeting platforms by relying on users' PII-based login data. For LiveRamp, the seed PII that builds the identifier is not expected to frequently change over time, potentially enabling participating actors to track the user over longer periods of time in comparison to current third-party cookies-based identifiers. Similarly, Google's Customer Match associates user's PII with the user's Google login, which is unlikely to change frequently.

In all three identification solutions under study, the user identifier is more persistent than a pseudo-anonymous third-party cookie identifier in the previous ecosystem. This might lead to potentially more persistent user tracking over time.

4.2.3 Opting-Out

All first-party identification architectures introduce technical improvements to the ability of users to exercise their rights, with dedicated opt-out mechanisms from personalised advertising that send opt-out signals to all actors in the bidding process about user's preferences (Asim, 2022; Google, n.d.-a; LiveRamp, 2022b; UnifiedID2, 2022). As opposed to current usage of third-party cookies, where trackers can easily escape or not provide functional opt-out options (Papadogiannakis et al., 2021), the first-party identification architectures enable an opt-out option from the start, by introducing pop-up windows that ask for user consent upon log-

4. Solutions such as Apple's hide my email can help periodically, but there is a limit to the obfuscation average users are capable of. Let alone when considering additional PIIs that are used to trace user identity over time.

ging in to a website.

Still, the likelihood that these opt-out mechanisms will allow users to fully realise their preferences remains questionable. Users may feel pressured against opting-out as the decision could mean losing access to publishers' content. In fact, the identification industry specifically states that the onus is on publishers to convince users to opt-in for targeted advertising in exchange for online content (Titone, 2021). Publishers are the ones who need to decide on the value exchange for users who do not wish to be targeted. They might prevent those users from viewing their content, making the designed opt-out mechanisms non-feasible across the different solutions. For walled garden products and other advertising platforms, users can refuse identification and targeting in the platforms' settings, but will have to be consistent. Researchers will have to verify that consent pop ups are not tricking users into accepting tracking terms they do not fully understand. Additionally, while Google's user choice and control documentation specifies that users can opt out from personalised ads that are served based on records of web browsing and app usage, it does not specify how users can manage their preferences when advertisers directly upload their PII into Google's Ads Data Manager and Customer Match for targeting.

4.2.4 Tracker transparency

Transparency regarding how third-party cookies are used for tracking and what data is collected about users has been insufficient (Fouad et al., 2022). The complex dynamics of tracking were not fully detailed and disclosures on privacy policies failed to describe observed tracking in practice (Papadogiannakis et al., 2021). In contrast, all first-party identification architectures promise to provide greater tracker transparency, but the extent to which users will be able to view records collected on their activity remains quite ambiguous. UID2.0 and LiveRamp's Ramp ID provide little insight into their promises of tracking transparency. They all provide 'transparency mechanisms,' but it is unclear how those mechanisms will provide up-to-date information to users on how they are being identified and by whom in a dynamic advertising delivery environment. Technically, in the complex settings of user identification and targeting for ads, it is almost impossible to provide users with accurate information on who is tracking them at any given moment. The first-party identification architectures under study are unclear on how exactly they will bridge this gap.

4.2.5 Sensitive targeting

All first-party identification architectures present significant concerns for sensitive

targeting, since they have the potential to make it easy and attractive for advertisers to target users based on sensitive attributes (e.g., race, religion, etc.) that are subject to legal or other limitations. Advertisers can persistently identify users based on their first-party customer data, and then ask targeting platforms to assemble 'look-alike audiences'[5]based on matched IDs. What that means is that an advertiser could use an attribute like race or religion to construct an audience, based on their first party-data, and then have an ad platform find those targets, effectively overriding the restrictions against this sort of discrimination that some ad platforms have made through their corporate policies. For instance, suppose an advertiser wants to target African-Americans on a social media platform that does not allow targeting by race. That advertiser could take a first-party dataset for which information about race is included (and for which a persistent identifier like an email address in known), then sort it into an audience defined by race, and then push that custom audience into the platform for activation without having to specifically select race-related targeting parameters from that platform. Further, the advertiser could request that the platform identify new targets, whose email or other IDs are not part of the advertiser first-party data set, via lookalike modeling (by other proxies such as postal code and income), which would be preprocessed according to a race variable. Such 'look-alike' audiences can now be targeted, enabling the targeting of African-Americans without specifically stating so to the targeting platforms.

Advertisers' capacities —and even incentives— to define and develop these proprietary audiences, before first-party user identifiers are encoded for targeting, represent a concerning feature of the systems under study here. Advertisers constantly purchase user data from data brokers to expand profiling of users beyond what passive surveillance of browser behaviour can provide. The emphasis on first-party data across the identification architectures is encouraging advertisers to further leverage their existing customer and sales data, amass additional data around those identifiable customers, and to find others with similar traits (LiveRamp, 2020). Compared to advertisers' efforts in the current third-party cookie-based ecosystem, advertisers are now investing more in technology that can match first-party customer data to other datasets (Vargas, 2022). This work is carried out via existing 'identity graph' and new 'clean room' technologies that allow advertisers to manage individual-level data and encode the data through an ID method of

5. 'Look-alike' audiences are potential advertising customers who resemble existing customers across many data attributes, and are therefore highly likely to share other data attributes with existing audiences. Advertisers / marketing agencies use look-alike audiences when they ask advertising platforms to target first-party identifiers of consumers that resemble existing consumers within the uploaded PII (Schneider, 2023).

choice; this provides a centralised system to merge online and offline identifiers into a consolidated profile to pair with purchased data from data brokers and activate selected audiences with multiple partners.

Within The Trade Desk Unified ID 2.0, for instance, the company mentions 'First-Party Relationships' capabilities, where an advertiser is able to upload first-party data to be encoded to the UID2.0 for activation across publisher sites (UnifiedID2, 2022). Similarly, LiveRamp offers advertisers the opportunity to 'onboard' their data where PII can be uploaded in order to be converted into RampIDs and organised by segment so that they can be activated in more than 500 different partner platforms (LiveRamp, 2022a). These capabilities resemble the primary service provided by Google's Customer Match previously described.

Importantly, there is no mechanism to verify how advertisers have segmented first-party data before importing into these systems, and the first-party identification architectures make sensitive population segmentation and bidding very attractive for advertisers. As detailed in Section 3, previous research has identified instances where advertiser-uploaded lists have violated the policies set by various ad sellers or intermediaries to prevent the targeting of users including categories such as race, religion, politics, sex life, or health (Wei et al., 2020b). We argue that first-party identification architectures place a significant amount of importance on leveraging first-party data that encourages advertisers to look for ways to link first-party and purchased data through ID solutions, allowing for targeting practices with little to no oversight by primary platforms (for example: Meta, 2021; Twitter, 2022).

TABLE 2 below summarises the privacy implications across first-party identification architectures and compares them to existing privacy concerns in third-party cookie-based identification settings.

PRIVACY CONCERN	UID 2.0	RAMPID	GOOGLE CUSTOMER MATCH
1 - CROSS-SITE TRACKING	ENABLED, BUT ONLY FOR THE UID OPERATOR. SSPS AND DSPS ONLY GET TO WORK WITH ENCRYPTED USER IDS AND EACH PUBLISHER GETS A DIFFERENT TOKEN..	SHARING & MATCHING OF IDS IS ENABLED ACROSS ADVERTISING PLATFORMS AND SITES FOR PARTICIPATING ACTORS.	ADVERTISERS ARE ABLE TO TARGET GOOGLE USERS ACROSS THIRD-PARTY PUBLISHERS THAT ARE PART OF THE GOOGLE DISPLAY NETWORK & ACROSS GOOGLE PRODUCTS.
2 - LONGITUDINAL TRACKING	PII-BASED USER IDENTIFIER IS CREATED, BASED ON USER LOGIN TO PUBLISHERS' WEBSITES. THIS IS MORE DETERMINISTIC THAN THIRD-PARTY COOKIE IDENTIFIERS AND CAN TRACK USERS OVER POTENTIALLY LONGER PERIODS OF TIME.	THE SEED FIRST-PARTY PII THAT BUILDS THE IDENTIFIER IS NOT EXPECTED TO FREQUENTLY CHANGE, POTENTIALLY ENABLING PARTICIPATING ACTORS TO TRACK THE USER OVER LONGER PERIODS OF TIME IN COMPARISON TO 3P COOKIES-BASED IDENTIFIERS.	GOOGLE'S CUSTOMER MATCH ASSOCIATES THE SEED FIRST-PARTY PII WITH THE USER'S GOOGLE LOGIN, WHICH IS UNLIKELY TO CHANGE FREQUENTLY, ENABLING TRACKING OVER LONG PERIODS OF TIME.
3 - OPTING-OUT	OPT-OUT MECHANISM HAS BEEN DESIGNED, BUT THE ONUS IS ON PUBLISHERS TO CONVINCE USERS TO OPT-IN IN EXCHANGE OF CONTENT, MAKING OPT-OUT CHOICES QUESTIONABLE IN PRACTICE.	OPT-OUT MECHANISM HAS BEEN DESIGNED. USERS CAN OPPOSE IDENTIFICATION AND TARGETING BY RAMPID. THE ONUS IS ON THE VARIOUS PARTICIPATING ACTORS FROM THE SELLER AND BUYER SIDES TO CONVINCE USERS TO OPT-IN IN EXCHANGE OF CONTENT.	OPT-OUT MECHANISM HAS BEEN DESIGNED. CONSUMERS OF DIFFERENT WALLED GARDEN PRODUCTS CAN CHOOSE TO OPT OUT FROM BEING TARGETED IN THEIR LOGGED-IN PRODUCTS.
4 - TRANSPARENCY IN TRACKING	SELF-GOVERNED TRANSPARENCY MECHANISM WAS DESIGNED, BUT THE INCLUDED INFORMATION IS UNCLEAR. IT IS TECHNICALLY ALMOST IMPOSSIBLE TO	SELF-GOVERNED TRANSPARENCY MECHANISM WAS DESIGNED, BUT THE INCLUDED INFORMATION IS UNCLEAR. IT IS TECHNICALLY ALMOST IMPOSSIBLE TO	SELF-GOVERNED TRANSPARENCY MECHANISM WAS DESIGNED, BUT THE INCLUDED INFORMATION IS UNCLEAR. IT IS TECHNICALLY ALMOST IMPOSSIBLE TO

	MONITOR FOR USERS ALL THE TRACKERS THAT COLLECT THEIR DATA AND TARGET THEM FOR ADS.	MONITOR FOR USERS ALL THE TRACKERS THAT COLLECT THEIR DATA AND TARGET THEM FOR ADS.	MONITOR FOR USERS ALL THE TRACKERS THAT COLLECT THEIR DATA AND TARGET THEM FOR ADS.
5- SENSITIVITY OF TARGETING	ENCOURAGE TARGETING BASED ON SENSITIVE CATEGORIES BY PAIRING ADVERTISER 1ST-PARTY & PURCHASED DATA AND THROUGH 'LOOK-ALIKE' AUDIENCE TARGETING.	ENCOURAGE TARGETING BASED ON SENSITIVE CATEGORIES BY PAIRING ADVERTISER 1ST-PARTY & PURCHASED DATA AND THROUGH 'LOOK-ALIKE' AUDIENCE TARGETING.	ENCOURAGE TARGETING BASED ON SENSITIVE CATEGORIES BY PAIRING ADVERTISER 1ST-PARTY & PURCHASED DATA AND THROUGH 'LOOK-ALIKE' AUDIENCE TARGETING.

5 - Discussion and conclusion

Our study details and evaluates privacy implications of the increasingly popular first-party identification architectures being used across the online advertising supply chain. We apply a five-part typology to show how the proposed first-party ID architectures maintain and, in some cases, even magnify the privacy concerns associated with third-party cookies. . Despite the marketing of those 'cookie-less' solutions as 'privacy-conscious' or 'privacy-first,' and the attempt of companies to distance their brands from the negative publicity surrounding online tracking, these first-party identification architectures reproduce AdTech's market logic of data-driven optimisation and efficiency through user identification & targeting, making very little progress, if any, for users' privacy.

The building of dossiers on users through cross-site tracking is enabled by all the solutions we examined. User identifiers are becoming more deterministic and persistent (often using personal data such as phone numbers and e-mail addresses), potentially enabling tracking over longer periods of time. The efficacy of the opt-out and self-governed tracking transparency mechanisms enabled by these new ID architectures remain questionable in practice, as it is heavily dependent upon proper self-implementation of governing arrangements.

The burden to obtain consent to first-party-based identifiers is on publishers, participating actors, and walled garden product managers, who need to convince users to accept the targeting in order to monetise their content. These first-party identification methods also incentivise advertisers to circumvent targeting policies

by providing greater means and incentives to target users based on sensitive categories. With the pivoting of the industry toward 'identity graphs,' 'clean rooms,' and first-party data, advertisers are now incentivised to target consumers based on the rich data profiles they hold on each user. In case sensitive categories like race or religion would sound appealing for advertisers, they can easily override existing targeting restrictions through 'look-alike' audience targeting schemes.

Ultimately, the business models and working practices of the advertising industry lead to a privacy-concerning ad landscape even without third-party cookie identifiers. The implementation of first-party identification architectures by the AdTech complex enables similar visibility on consumers, longer consumer tracking, and the assembling of more sensitive consumer targeting, with user agency and tracker transparency still very questionable. AdTech trackers can still learn about users' demographics data, interest data, intent data (intention to purchase), and measurement data (attribution), maintaining their data-driven business model that does not go hand-in-hand with meaningful privacy.

Despite their apparent resignation to pro-privacy rhetoric, advertisers, advertising technology companies and the publishers who sell access to audiences have not abandoned their commitments to data-driven efficiency and optimisation (Veale, 2022); they still want to recognise, as precisely as possible, the probable behaviours of consumers and the expected returns on their advertising investments (McGuigan, 2023). Creators of first-party identification architectures push for targeting of users based on their first-party data, creating an advertising landscape where only they can fully capitalise on the shift away from third-party cookies. The way those companies market 'first-party privacy-preserving Ad Tech' appear like a smokescreen to the real targeting and identification solutions they are pushing and advocating for.

To diverge from this path, we call for the use of user identities in the technical sense to be constrained to protect identity construction in the sense of selfhood (Hildebrandt, 2013). Such a reform would require a structural change in how online advertising currently works. Instead of allowing advertising actors to collect and purchase more users' data, regulators would have to limit the attributes that advertisers can collect on users, and constrain their ability to draw inferences outside of the context in which user data were collected. That should also include limits on advertisers' ability to assemble targeting criteria.

The history of technological development for the delivery of ads has proven time and again that change will not come from users' ability to opt-out and understand

the level of tracking they face. Nor will it emerge from advertisers' concerns about users' privacy. Change has to come from public-interest regulators who can fully appreciate the structural gaps that enable privacy abuse in the online advertising ecosystem.

References

Alaimo, C. (2022). From people to objects: The digital transformation of fields. *Organization Studies*, *43*(7), 1091–1114. https://doi.org/10.1177/01708406211030654

Beauvisage, T., Beuscart, J.-S., Coavoux, S., & Mellet, K. (2024). How online advertising targets consumers: The uses of categories and algorithmic tools by audience planners. *New Media & Society*, *26*(10), 6098–6119. https://doi.org/10.1177/14614448221146174

Choi, H., Mela, C. F., Balseiro, S. R., & Leary, A. (2020). Online display advertising markets: A literature review and future directions. *Information Systems Research*, *31*(2), 556–575. https://doi.org/10.1287/isre.2019.0902

Crain, M. (2021). *Profit over privacy: How surveillance advertising conquered the Internet*. Univ Of Minnesota Press.

Eng, A. (2024, June 6). *Key trends in digital advertising: Adapting to the new privacy norm*. IAB. https://www.iab.com/blog/2024-key-trends-in-digital-advertising/

European Commission. Directorate General for Communications Networks, Content and Technology. (2023). *Study on the impact of recent developments in digital advertising on privacy, publishers and advertisers: Final report*. Publications Office. https://data.europa.eu/doi/10.2759/294673

Fouad, I., Bielova, N., Legout, A., & Sarafijanovic-Djukic, N. (2020). *Missed by filter lists: Detecting unknown third-party trackers with invisible pixels* (No. arXiv:1812.01514). arXiv. https://doi.org/10.48550/arXiv.1812.01514

Fouad, I., Santos, C., Legout, A., & Bielova, N. (2022). My cookie is a phoenix: Detection, measurement, and lawfulness of cookie respawning with browser fingerprinting. *Proceedings on Privacy Enhancing Technologies*. https://petsymposium.org/popets/2022/popets-2022-0063.php

Fourcade, M., & Healy, K. (2017). Seeing like a market. *Socio-Economic Review*, *15*(1), 9–29. https://doi.org/10.1093/ser/mww033

Google. (n.d.-a). *About the customer matching process*. Google Ads Help. https://support.google.com/google-ads/answer/7474263

Google. (n.d.-b). *Create a customer list*. Google Ads Help. https://support.google.com/google-ads/answer/6276125

Google. (n.d.-c). *Format your customer data file*. Google Ads Help. https://support.google.com/google-ads/answer/7659867?sjid=13277765107387728213-NA#zippy=

Google. (n.d.-d). *How Google uses Customer Match data*. Google Ads Help. https://support.google.com/google-ads/answer/6334160?sjid=13277765107387728213-NA

Google. (n.d.-e). *Upload Customer Match data*. Google Ads Help. https://support.google.com/google-ads/answer/10589050?hl=en

Google. (2024). *About Customer Match*. Google Ads Help. https://support.google.com/google-ads/answer/6379332?sjid=10490414924746101 56-NA

Graßl, P., Schraffenberger, H., Zuiderveen Borgesius, F., & Buijzen, M. (2021). Dark and bright patterns in cookie consent requests. *Journal of Digital Social Research*, *3*(1), 1–38. https://doi.org/10.33621/jdsr.v3i1.54

Hercher, J. (2019, July 19). *The companies challenging LiveRamp's supremacy in data onboarding*. Adexchange. https://www.adexchanger.com/data-exchanges/the-companies-challenging-liveramps-supremacy-in-data-onboarding/

Hercher, J. (2024, August 9). *The Trade Desk says UID2 has now reached 'critical mass'*. Digiday. https://digiday.com/media/some-publishers-are-starting-to-see-revenue-lift-from-alternative-ids/

IAB. (2024). *IAB/PWC advertising revenue report*. IAB. https://www.iab.com/insights/internet-advertising-revenue-report-2024/

IAB Canada. (2021). *A guide to authenticated audiences & universal IDs*. IAB Canada. https://iabcanada.com/wp-content/uploads/2021/06/IAB-Canada-UID-June-2021.pdf

Janardhan, K. (2023, October 11). *Simplifying the management of your first-party data*. Google Blog. https://blog.google/products/ads-commerce/simplifying-the-management-of-your-first-party-data/

Kant, T. (2021). Identity, advertising, and algorithmic targeting: Or how (not) to target your "ideal user". *MIT Case Studies in Social and Ethical Responsibilities of Computing*. https://doi.org/10.21428/2c646de5.929a7db6

Karaj, A., Macbeth, S., Berson, R., & Pujol, J. M. (2019). *WhoTracks .Me: Shedding light on the opaque world of online tracking* (No. arXiv:1804.08959). arXiv. https://doi.org/10.48550/arXiv.1804.08959

Libert, T., & Binns, R. (2019). Good news for people who love bad news: Centralization, privacy, and transparency on US news sites. *Proceedings of the 10th ACM Conference on Web Science*, 155–164. https://doi.org/10.1145/3292522.3326019

Macbeth, S. (2017). *Tracking the trackers: Analysing the global tracking landscape with GhostRank* [Technical Report]. Cliqz GmbH. https://cdn.cliqz.com/wp-content/uploads/2017/12/Ghostery_Study_-_Tracking_the_Trackers.pdf

Martin, K. (2015). Privacy notices as Tabula Rasa: An empirical investigation into how complying with a privacy notice is related to meeting privacy expectations online. *Journal of Public Policy & Marketing*, *34*(2), 210–227. https://doi.org/10.1509/jppm.14.139

McGuigan, L. (2023). *Selling the American people: Advertising, optimization, and the origins of Adtech*. The MIT Press.

Meta. (2021, November 9). *Removing certain ad targeting options and expanding our ad controls*. Facebook. https://www.facebook.com/business/news/removing-certain-ad-targeting-options-and-expanding-our-ad-controls

Papadogiannakis, E., Papadopoulos, P., Kourtellis, N., & Markatos, E. P. (2021). User tracking in the post-cookie era: How websites bypass GDPR consent to track users. *Proceedings of the Web Conference 2021*, 2130–2141. https://doi.org/10.1145/3442381.3450056

Papadopoulos, P., Kourtellis, N., & Markatos, E. (2019). Cookie synchronization: Everything you always wanted to know but were afraid to ask. *The World Wide Web Conference*, 1432–1442. https://doi.org/10.1145/3308558.3313542

Samarasinghe, N., & Mannan, M. (2019). Towards a global perspective on web tracking. *Computers & Security*, *87*, 101569. https://doi.org/10.1016/j.cose.2019.101569

Sherman, J. (2021). *Data brokers and sensitive data on US individuals: Threats to American civil rights, national security, and democracy*. Duke University Sanford School of Public Policy. https://techpolicy.sanford.duke.edu/wp-content/uploads/sites/4/2021/08/Data-Brokers-and-Sensitive-Data-on-US-Individuals-Sherman-2021.pdf

Shields, R. (2023, June 5). *Amid a dearth of ad tech M&A, LiveRamp fielded inbound inquiries*. Digiday. https://digiday.com/marketing/amid-a-dearth-of-ad-tech-ma-liveramp-fielded-inbound-inquiries-over-a-potential-sale/

Sivan-Sevilla, I., & Poudel, P. (2024). *Web privacy based on contextual integrity: Measuring the collapse of online contexts*. arXiv. https://doi.org/10.48550/ARXIV.2412.16246

Solomos, K., Ilia, P., Ioannidis, S., & Kourtellis, N. (2020). *Clash of the trackers: Measuring the evolution of the online tracking ecosystem* (No. arXiv:1907.12860). arXiv. https://doi.org/10.48550/arXiv.1907.12860

Statista Research Department. (2024). *US online advertising revenue 2023*. Statista.Com. https://www.statista.com/statistics/183816/us-online-advertising-revenue-since-2000/

Tabisz, J. (2023, May 18). *Brands make big bet on first-party data as third-party cookie fades away*. Digiday. https://digiday.com/marketing/digiday-research-brands-make-big-bet-on-first-party-data-as-third-party-cookie-fades-away/

The Trade Desk. (2021, February 21). *What the tech is unified ID 2.0?* The Current. https://www.thetradedesk.com/us/news/what-the-tech-is-unified-id-2-0

Titone, T. (2021, May 3). *Unified ID 2.0 explained*. Ad Tech Explained. https://adtechexplained.com/unified-id-2-0-explained/

Trevisan, M., Traverso, S., Bassi, E., & Mellia, M. (2019). 4 years of EU Cookie Law: Results and lessons learned. *Proceedings on Privacy Enhancing Technologies*, *2019*(2), 126–145. https://doi.org/10.2478/popets-2019-0023

Turow, J. (2011). *The daily you: How the new advertising industry is defining your identity and your worth*. Yale University Press.

Twitter. (2022). *Targeting of sensitive categories*. X Business. https://business.twitter.com/en/help/ads-policies/campaign-considerations/targeting-of-sensitive-categories.html

Vargas, A. (2022, April 12). *Goodway group stitches together identity graph to complement brands' first-party data*. Adexchange. https://www.adexchanger.com/agencies/goodway-group-stitches-together-identity-graph-to-complement-brands-first-party-data/

Veale, M. (2022, February 25). *Future of online advertising: Adtech's new clothes might redefine privacy more than they reform profiling*. NetzPolitik.Org. https://netzpolitik.org/2022/future-of-online-advertising-adtechs-new-clothes-might-redefine-privacy-more-than-they-reform-profiling-cookies-meta-mozilla-apple-google/

Veale, M., & Borgesius, F. Z. (2022). Adtech and real-time bidding under European Data Protection

Law. *German Law Journal*, *22*, 226–256.

Wachter, S. (2020). Affinity profiling and discrimination by association in online behavioural advertising. *Berkeley Technology Law Journal*, *35*(2). https://doi.org/10.2139/ssrn.3388639

Wei, M., Stamos, M., Veys, S., Reitinger, N., Goodman, J., Herman, M., Filipczuk, D., Weinshel, B., Mazurek, M. L., & Ur, B. (2020). What Twitter knows: Characterizing ad targeting practices, user perceptions, and ad explanations through users' own Twitter data. *29th USENIX Security Symposium (USENIX Security 20)*, 145–162.

Yang, Z., & Yue, C. (2020). A comparative measurement study of web tracking on mobile and desktop environments. *Proceedings on Privacy Enhancing Technologies*, *2020*(2), 24–44. https://doi.org/10.2478/popets-2020-0016

Yuen, M. (2024, August 2). *Guide to Google for marketers and advertisers*. Emarketer. https://www.emarketer.com/learningcenter/guides/guide-google/

Zwick, D., & Dholakia, N. (2004). Whose identity is it anyway? Consumer representation in the age of database marketing. *Journal of Macromarketing*, *24*(1), 31–43. https://doi.org/10.1177/0276146704263920

Published by

Alexander von Humboldt Institute for Internet and Society

CAIS Research for the Digital Age

in cooperation with

CREATe

cnrs centre internet et societe

R&I IN3 Internet interdisciplinary Institute Universitat Oberta de Catalunya

University of Tartu Johan Skytte Institute of Political Studies