The advent of privacy-centric digital advertising: Tracing privacy-enhancing technology adoption

Garrett A. Johnson & Nico Neumann*

March 21, 2024, *Work in progress*

Abstract

In response to privacy regulation (e.g., GDPR, CCPA), the digital ad industry is adopting privacy-enhancing technologies to deliver relevant advertising while improving consumer privacy. In particular, Google proposes to sunset third-party cookies and replace them an ensemble of new technologies it calls the "Privacy Sandbox." However, regulators are scrutinizing the effectiveness and competitive consequences of these technologies. In this project, we measure the adoption of Google's Privacy Sandbox for a panel of almost 60,000 top commercial websites. We show that current adoption varies by technology: at least 1-5% of sites use the Protected Audience API whereas at least 36% use the Topics API in 2023. Moreover, we measure associated market shares for adtech vendors (e.g., Criteo, Google, RTB House). We share our updated findings on a public dashboard (`app.sincera.io/privacysandbox`) to aid stakeholders like academics, firms, journalists, and regulators.

---

*Johnson: Questrom School of Business, Boston University <garjoh@bu.edu>; Neumann: Melbourne Business School, University of Melbourne <n.neumann@mbs.edu>. This is work in progress: data collection and analysis are ongoing. We gratefully acknowledge the financial support of the Digital Business Institute at Boston University Questrom School of Business, the Rafik B. Hariri Institute for Computing and Computational Science & Engineering at Boston University, and the Program on Economics & Privacy at George Mason University. We thank Zhengrong Gu, and Alisha Ukani as well as audiences at the 2024 ACM Symposium on Computer Science and Law, Consumer Financial Protection Bureau, George Mason University, and Topics W3C meeting for helpful comments. We thank Aman Ahmed, Yufan Biao, and Nakul Shahdadpuri for their research assistance. We thank Ian Meyers and Michael O'Sullivan at Sincera for data assistance. *Disclosures*: Prof. Johnson serves as an outside expert for the Competition and Markets Authority (CMA) in the United Kingdom for its Privacy Sandbox case. This research was conducted in his personal capacity and does not reflect the views of the CMA.

1 Introduction

Digital revolutionized advertising by transforming how ads could be targeted, measured, and optimized. Cross-site or cross-app user identifiers like cookie and mobile identifiers underpin these capabilities. In particular, advertisers gained the ability to target users based on their interests as inferred from their past-browsing data (Goldfarb, 2013). Advertisers could now measure ad effectiveness by connecting data on user ad views to subsequent user actions like site visits and purchases. Finally, advertisers gained the ability to optimize ad campaigns in real time by bidding for ads according to the predicted value of such user actions or ad clicks (Wernerfelt et al., 2022). These new capabilities created value for advertisers, which was shared with ad intermediaries and websites. In particular, some research shows that ad prices approximately double when ad impressions include a cookie identifier (Johnson et al., 2020; Ravichandran & Korula, 2019).

At the same time, these online ad practices provoked privacy concerns among consumers and regulators. Europe in particular sought to limit these practices with successive regulations including the e-Privacy Directive, the General Data Protection Regulation (GDPR), and the Digital Services Act. Many browsers—including Firefox, Safari, and Brave—introduced technological means to block cross-site identifiers. In April 2021, Apple enforced a strict user consent policy for sharing cross-app identifiers on its iOS mobile platform (Sokol & Zhu, 2021). Google committed to deprecating third-party cookies on its Chrome browser in 2024 and removing the ad identifier on the Android mobile platform soon after.

To resolve this seemingly fundamental trade-off between privacy and value creation in this industry, some technology firms seek to apply privacy-enhancing technologies (PETs) to solve the industry's key use cases. These firms intend to rebuild digital advertising on a technology foundation that guarantees a level of privacy to consumers. In particular, Google's is leading this effort to replace cross-site/app identifiers with a suite of proposed technologies, which it collectively calls the "Privacy Sandbox" (Google, 2022).

Google's "Privacy Sandbox" consists of multiple proposals that aim to preserve many of the benefits of cross-site identity in programmatic advertising while offering superior privacy protection to the user. These include technologies for ad targeting (Protected Audience API & Topics API), measurement (Attribution Reporting), and fraud detection (Private State Tokens). These technologies use several privacy-enhancing approaches including data aggregation, k-anonymity, differential privacy, and on-device computation. These proposals and similar PET applications have the potential to revolutionize how practitioners and researchers use digital advertising (Johnson et al., 2022).

In this paper, we track the real world adoption of Privacy Sandbox technologies by websites and adtech vendors. We partnered with Sincera, a technology startup company, to develop technology that detects the adoption of Privacy Sandbox by websites. We scan a panel of over 58,000 top commercial websites and identify which websites use which Privacy Sandbox technologies. We also intercept data that websites share with these technologies to identify their adtech vendor partners and learn how the sites and vendors use these technologies. We focus on two key advertising technologies within the Privacy Sandbox: the Protected Audience and Topics APIs.

We provide *conservative underestimates* of Privacy Sandbox adoption for several reasons. First, we do not detect Protected Audience API use by Google's adtech until February 2024 for technical reasons. Second, we do not observe vendors who block API calls originating from cloud data centers. Third, we exclusively measure JavaScript calls to Privacy Sandbox APIs, but omit vendors access the APIs through other means: in particular, we omit HTTP header calls to the Topics API. Fourth, we omit vendors that exclusively call the APIs for users that were randomly selected to participate in Google's Privacy Sandbox experiment. Fifth, we miss API calls by vendors that require active user consent.

By the end of 2023, we observe moderate adoption of these technologies by websites and by adtech vendors. We observe that the web sites that use the Protected Audience API to show targeted ads to users in 2023 peaked at 4.7%. We also see that the share of sites that use Protected Audience API to show targeted ads on their site is at least 1.1% of sites at the end of 2023 (as indicated by site's header bidding settings). We observe that vendors request a user's behavioral interest segment using the Topics API on 36.2% of websites. 1.0% of websites signal that they opt out from participating in Topics API, which is down from 2.7%-3.9% in July 2022. We continue to study adoption along the path to third-party cookie deprecation. Google made Privacy Sandbox technologies generally available on Chrome browsers in September 2023, deprecated cookies for 1% of users in January 2024 as part of a testing phase, and proposes to fully deprecate cookies in the latter half of 2024.

As part of this research, we are building a public dashboard that tracks website adoption of Privacy Sandbox technologies: `sincera.io/privacy-sandbox`. In doing so, we intend to improve transparency of these novel technologies. This dashboard will provide updated statistics on the adoption rate over time as well as the market share of key adtech vendors. The dashboard can help academics, industry participants, journalists, and regulators better understand these new technologies. The dashboard can thereby inform the public discussion on privacy-centric advertising, since the complexity of these technologies may engender consumer skepticism.[1]

[1]See e.g., `https://mobiledevmemo.com/the-privacy-sandbox-paradox/`.

Our study contributes to several domains. We contribute to the marketing and economics literature on privacy (Acquisti et al. 2016; Bleier et al. 2020; Goldfarb & Que 2023; Goldfarb & Tucker 2024) and privacy in online advertising in particular (Tucker, 2012; Choi & Jerath, 2022). Many scholars investigate the tension between privacy and the value of user identifiers to online advertising: both third-party cookies (see e.g., Goldfarb & Tucker, 2011; Johnson et al., 2020; Miller & Skiera, 2023) and mobile ad identifiers (Cecere & Lemaire, 2023; Kraft et al., 2023). This literature explores how industry uses contextual, ad measurement, and user behavioral data to optimize advertising (see e.g., Ada et al., 2022; Farahat & Bailey, 2012; Rafieian & Yoganarasimhan, 2021; Wernerfelt et al., 2022). Johnson et al. (2022) describe the privacy-centric advertising technologies proposed by Apple and Google and discuss their implications for research. Alcobendas et al. (2023) predict the impact of Privacy Sandbox on advertisers and publishers using a structural approach. Jerath & Miller (2023) compare consumer responses to Privacy Sandbox to alternative ad targeting approaches. We instead investigate how the industry adopts Privacy Sandbox in practice.

We contribute to the economic literature on the diffusion and consequence of new technologies diffusion (e.g., Forman et al., 2005; Zolas et al., 2021). Our setting is notable in that online advertising may be the first industry to experiment with technologies where PETs serve as the foundation. PETs promise to balance privacy concerns with the value created by data for the economy. Regulators are interested in the potential role of PETs in privacy compliance (ICO 2022; OPCC 2017), though the economics literature on PETs is nascent (Acquisti, 2024; Tucker, 2024). For instance, several marketing and information systems scholars examine how consumer data can be transformed to improve privacy while preserving data utility (Anand & Lee, 2023; Li et al., 2022; Schneider et al., 2018).

Both regulators and scholars are concerned about the consequences of Privacy Sandbox for competition. In general, industrial standards affect competition (David & Greenstein, 1990) and PETs themselves may limit competition. Privacy Sandbox raises antitrust concerns (Geradin et al., 2021a) and faces scrutiny from the British Competition and Markets Authority (CMA).[2] Indeed, researchers have generally found that privacy restrictions harmed competition in online advertising and adjacent industries including the GDPR (Janssen et al., 2022; Johnson et al., 2023b; Peukert et al., 2022), COPPA (Johnson et al., 2023a; Kircher & Foerderer, 2024), and Apple's privacy restrictions on iOS (Aridor & Che, 2024; Li & Tsai, 2023). Toward this, we describe the structure the adtech vendor market for vendors employing Privacy Sandbox tools. Moreover, identifying Privacy Sandbox adopters can help regulators evaluate the effectiveness of Privacy Sandbox given the challenge of limited marketplace adoption and testing.

[2] `https://www.gov.uk/cma-cases/investigation-into-googles-privacy-sandbox-browser-changes`.

We also contribute to the web measurement literature in computer science that studies websites' use of advertising vendors and resulting consequences for user privacy. Until now, a browser—upon opening a website—interacts with potentially dozens of third-party domains of various vendors that perform services for that website. Computer scientists document websites' use of vendors and the technologies they employ to identify users (e.g., Englehardt & Narayanan 2016; Lerner et al. 2016; Libert 2015; Nikiforakis et al. 2013). This data is notable because it affords transparency into both data sharing between firms as well as website-vendor networks. Public databases of this kind include `httparchive.org` and WhoTracksMe (Karaj et al., 2018). More recently, economists and marketing researchers used such data to study the economic consequences of the GDPR (Lefrere et al., 2022; Lukic et al., 2023; Johnson et al., 2023c; Peukert et al., 2022). However, privacy regulation and platform changes are forcing industry practices to evolve, though newer approaches may reduce transparency. To address this, we contribute a novel measurement approach to detect website and vendor use of privacy-centric advertising technologies. Our work is closest to Rumiński et al. (2022), which provides a snapshot of Protected Audience API adoption in August 2022. Chrome provides monthly public statistics on the share of page loads that use HTML or Javascript features related to Privacy Sandbox,[3] whereas we construct more granular website- and vendor-level data.

We contribute to the economics and marketing literature that examines the fast-evolving online advertising industry and its challenges (see e.g., Gordon et al., 2021). Privacy Sandbox usage data promises to increase transparency in this industry, which can reveal how advertisers and publishers use new tools to target and sell ads. For instance, Neumann et al. (2019) and Neumann et al. (2023) show that cookie-based behavioral data had quality issues, but Protected Audience API and Topics API may improve both data quality and transparency. Many scholars examine the real-time auctioning of online ads (Choi et al., 2020). For instance, Akbarpour & Li (2020) highlight incentive issues for online ad auction rules. We illuminate websites' choices of vendor partners and auction structure within Protected Audience API.

The remainder of this paper is organized as follows. Section 2 describes Privacy Sandbox and the specific technologies that we measure. Section 3 describes our data and measurement approach. Section 4 provides our Privacy Sandbox adoption results. Section 5 concludes.

[3]These statistics are collated on a dedicated dashboard at `https://pscs.glitch.me/`. For instance, Topics API (specifically "TopicsAPI_BrowsingTopics_Method") use can be viewed at `https://chromestatus.com/metrics/feature/timeline/popularity/4182`.

2 Background: Google's Privacy Sandbox

Google Chrome is one of the last major browser to support third-party cookies. Due to privacy concerns, Safari blocked third-party cookies by default in 2018, followed by Firefox in 2019. Third-party cookies contain pseudonymous identifiers that facilitate identifying users across websites. This is valuable because it enables online behavioral advertising: i.e., targeting users based on the sites users browse. Behavioral advertising includes segmenting users by their interest as inferred by their browsing behavior as well as retargeting: identifying users who interact with an advertiser and advertising to them on unrelated websites. Third-party cookies also facilitate ad effectiveness measurement by connecting user ad views to ad clicks as well as user conversions (e.g., purchases, store lookups) on the advertiser's site. Adtech vendors work with a network of websites to place their third-party cookies on user browsers. Adtech vendors combine real-time, cross-site user interest and ad effectiveness data to optimize their advertisers campaign objectives. Beyond advertising, third-party cookies are also used for user authentication (i.e., single sign-on), cross-site content personalization, and fraud prevention.

Google (2022) is developing replacement technologies for third-party cookies under its "Privacy Sandbox" umbrella. Google states that it seeks to improve user privacy while preserving the benefits of personalized advertising like funding websites and apps.[4] Moreover, Google argues that merely blocking third-party cookies led advertisers to instead employ more covert forms of tracking like fingerprinting.[5] The Privacy Sandbox consists of multiple technologies designed to serve specific use cases. Advertising use cases include targeting (Protected Audience API and Topics API), measurement (Attribution Reporting API), and fraud prevention (Private State Tokens API) as well as user authentication (Federated Credential Management API). These proposals employ a variety of PETs including on-device computing, differential privacy, and zero-knowledge proofs (see, e.g., ICO 2022 for an introduction to PETs). Privacy Sandbox also includes technologies that improve privacy by limiting information flow across websites (e.g., CHIPS API, Fenced Frames API) and reduce covert tracking (e.g., User-Agent Reduction, Network State Partitioning).

Google proposed Privacy Sandbox in 2019. Though many core ideas remain from these initial proposals, Google's Privacy Sandbox evolved over time in response to feedback as some elements were shelved (e.g., FLoC: Federated Learning of Cohorts) while others were renamed (e.g., TURTLEDOVE became FLEDGE became Protected Audience API). Google released trials of several APIs in 2022, though this was restricted to a small

[4] https://blog.google/products/chrome/building-a-more-private-web/.

[5] https://privacysandbox.com/news/working-together-to-build-a-more-private-internet/.

fraction of users. Google made Privacy Sandbox generally available for Google Chrome user in September 2023.[6] On January 4, 2024, Google deprecated third-party cookies for 1% of Chrome users. These users were randomly selected as part of an industry-wide experiment in consultation with the British CMA to evaluate Privacy Sandbox.[7] Pending CMA approval, Google intends to phase out third-party cookies in the second half of 2024.

Academics from multiple disciplines have interrogated the Privacy Sandbox. In marketing and economics, Alcobendas et al. (2023) use a structural model to simulate privacy-related restrictions to online advertising market. Alcobendas et al. (2023) suggest that Privacy Sandbox mitigates the loss in publisher revenue and advertiser surplus associated with banning third-party cookies. However, Alcobendas et al. (2023) warn that Privacy Sandbox could hurt competition by strengthening the relative position of firms with large, first-party data relationships with users. Cooper et al. (2023) find that surveyed consumers tend to favor Privacy Sandbox over identity-based replacements to third-party cookies. Similarly, Jerath & Miller (2023) find that consumer privacy perceptions show modest improvement in Privacy Sandbox-related targeting scenarios relative to traditional behavioral targeting. Martin et al. (2023) and McGuigan et al. (2024) provide a potential explanation: despite objective improvements in limiting data access and data minimization, behavioral-targeted advertising may fundamentally remain counter to some consumers' privacy expectations. Nevertheless, Google researchers find that users report that topic-based personalization improves their perception of privacy and feeling of control relative to third-party cookies (Lachner et al., 2023).[8] Law scholars grapple with the competitive consequences of Privacy Sandbox (Geradin et al., 2021b; Nottingham, 2021), its governance (Olejnik, 2023), as well as its fit with European privacy law (Olejnik, 2024).

We focus on Google's Privacy Sandbox because it is generally available on the web. Other tech firms have proposed their own PET-based approaches to online advertising—e.g., Microsoft's Ad Selection API (formerly PARAKEET) and Firefox and Meta's joint Interoperable Private Attribution proposal—though these are in the development stage. Notably, Microsoft's Ad Selection API builds on Google's Protected Audience API, and enters testing in the second half of 2024.[9] Apple's launched a third-party cookie replacement technology—Private Click Measurement—though this exclusively addresses ad measurement. In particular, Apple's Private Click Measurement generates ad mea-

[6] `https://privacysandbox.com/news/privacy-sandbox-for-the-web-reaches-general-availability`.

[7] `https://developers.google.com/privacy-sandbox/setup/web/chrome-facilitated-testing`.

[8] Note that concerned users can opt out of Privacy Sandbox at the technology level or instead block specific Topic API segments and specific Protected Audience API advertiser domains. User controls can be accessed within Chrome's browser settings or by navigating to `chrome://settings/adPrivacy`.

[9] `https://blogs.windows.com/msedgedev/2024/03/05/new-privacy-preserving-ads-api/`.

surement reports without websites or advertisers needing to take dedicated actions like calling this function. While this simplifies industry use, this prevents researchers from measuring adoption.

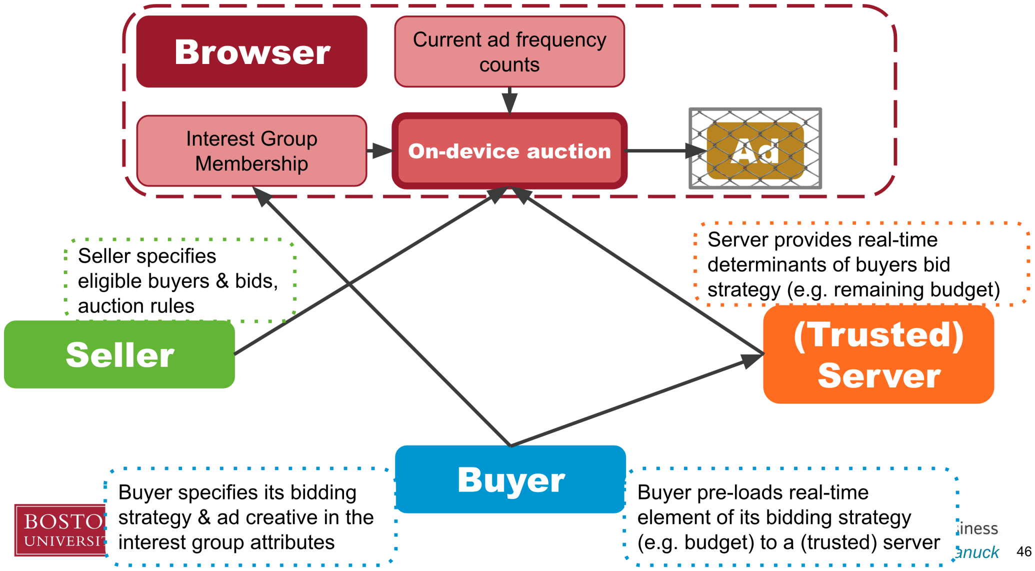
We study the adoption of two Privacy Sandbox technologies that we discuss below: Protected Audience API and Topics API.

2.1 Protected Audience API

The Protected Audience API (PA API) rearchitects online ad transactions to instead take place on a user's device. The canonical PA API use case is retargeting. Currently, retargeting allows an online advertiser and publisher to learn that a user visited both sites, which creates some privacy risk. With PA API, the online retailer tells the browser to add site visitors to an "interest group" for ad serving. This interest group encodes the static elements of the advertiser's bidding preferences as well as a link to the ad creative. When the user visits the publisher site, the site instructs the browser on the ad auction rules and the browser runs an auction on the publisher's behalf using all the user's interest groups stored on the browser. Figure 1 summarizes how PA API works to choose ads for users. PA API is appealing from a research perspective as we can use this to gain insight on how certain advertisers assign visitors to ad targeting groups. On the publisher side, we see which publishers use PA API to sell ads, and can learn about the publisher's ad selling preferences. In Section 4.1, we examine the share of sites that add users to interest groups and the share of sites that initiate a PA API auction as well as the market shares of their corresponding ad tech vendors.

Adtech vendors RTB House and Criteo provide an early look at PA API testing and adoption in 2022. RTB house worked with its advertisers to experiment with retargeting campaigns via PA API in the summer of 2022 (Rumiński et al., 2022). At the time, Privacy Sandbox was enabled for less than 1% of Chrome beta users. During that time, RTB House added 1.2 million users globally to PA interest groups. RTB House showed over 7 million PA impressions from nearly 400 advertisers on 4,172 advertiser domains. RTB House also collected data on interest group creation on 13,354 sites during one week in August 2022. Rumiński et al. (2022) observed 18,207 created interest groups and their associated ad tech vendors: Google (53.6%), RTB House (33.4%), and Criteo (12.9%). By the end of 2022, Criteo was adding about 80 million users to interest groups per day across 14,000 websites when Privacy Sandbox was available on 5% of Chrome users (Höring, 2022). In the last quarter of 2022, Criteo was receiving about 10 million PA bid requests daily. Criteo and RTB House flagged low adoption by supply-side platforms in the ad tech ecosystem as a challenge for PA adoption and use. Other industry testers flagged latency

Figure 1: Visual explanation of targeted ad serving with Protected Audience API



and debugging as other issues.[10]

2.2 Topics API

Topics API groups users into a few interest segments—chosen from 469 segments, like "tennis" or "cats"[11]—based on the domain names of sites that the user browsed in the past. Topics API only classifies users based on the domains they browse and not the full URL address or the page content. Google, rather than the domain owners, assigns domains to segments by manually labeling 50,000 top domains and then employing a classifier model to other website hostnames.[12] Google further splits the Topics API segments by their top-level category into those that it deems to be more or less commercially relevant: e.g., "Finance" and "Shopping" versus "News" and "Online Communities."[13] The browser selects the user's top five most visited categories after first sorting for commercial relevance. Google added this emphasis on commercial relevance in response to industry feedback (see e.g., Selman, 2022). For each of the three most recent weeks, the Topics API returns either one of these top five categories (with equal probability) or a randomly selected category with 5% probability. Thus, Topics API returns at most three

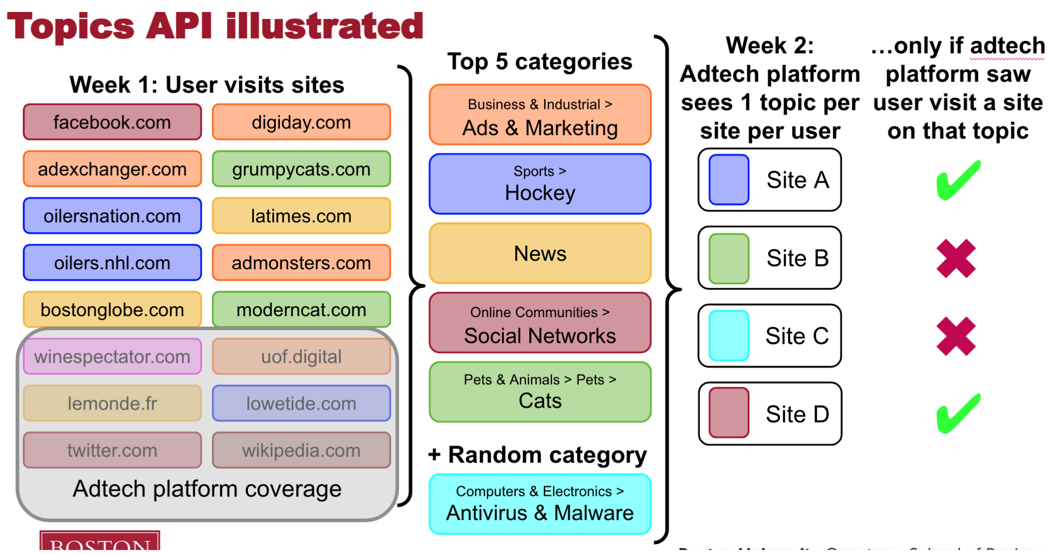
[10] See, e.g., https://github.com/grupawp/PAapi and https://www.adexchanger.com/the-sell-sider/privacy-sandboxs-latency-issues-will-cost-publishers.

[11] Topics API originally included 350 segments, but version two of its taxonomy expanded to 469 segments. For the full taxonomy, see https://github.com/patcg-individual-drafts/topics/blob/main/taxonomy_v2.md.

[12] https://developers.google.com/privacy-sandbox/relevance/topics/topic-classification.

[13] https://developers.google.com/privacy-sandbox/blog/topics-enhancements.

Figure 2: Visual explanation of how Topics API captures and shares user behavior-based interest segments



segments per user. Advertisers can use this segment information to prospect for potential customers or to improve their user match-value predictions when bidding on ads.

Figure 2 visualizes how Topics API works. Adtech vendors that are present on a participating website can call Topics API on the user's browser to view the user's interest segments. However, ad vendors only obtain segment information for users that they have seen on domains relevant to that segment. This creates an incentive for ad vendors to obtain wide coverage across different websites. The Topics API randomizes the segments that it returns at the level of the vendor and originating domain. This makes the Topics API segments less useful for tracking users (i.e., fingerprinting) since vendors receive different segments for the same user on different domains (Alvim et al., 2023; Carey et al., 2023; Epasto et al., 2022). However, Firefox's Thomson (2023) critiques both the privacy properties of Topics API as well as its utility for advertisers. Note that websites can also choose to opt out of participating in Topics and having their data used for this purpose. In Section 4.2, we report data on vendors that make calls to the Topics API, the share of sites where these calls are present, and the share of sites opt out of Topics.

In testing, Google (2023) found that Topics API performs almost as well as its third-party cookie-based behavioral segments. However, Criteo found that its cookie-based, in-market segments performed much better than Topics API in terms of those users performing relevant actions in a given category (Selman, 2022). RTB House reported that it only saw Topics API segments in 0.15% of bid requests from Chrome users at the end of 2024 (Król, 2023). Both Criteo and RTB House report the top Topics API segments they observe in user data. However, this list is dominated by generic categories like "News" and

"Arts & Entertainment" as the data precede changes to Topics API: the updated taxonomy as well as pre-sorting segments by commercial use. Criteo and Xandr report that the Topics API site classifier broadly agrees with their respective in-house classifiers (Quéré, 2022; Selman, 2022).

3 Data

We collect data on Privacy Sandbox technology adoption and use over time by a panel of over 55,000 top websites. Below, we describe our data collection methodology and then describe our sample of websites.

3.1 Data collection

We work with Sincera (`sincera.io`), an American start-up company, to collect data on websites' adoption of Privacy Sandbox technologies. Sincera offers a paid service that scans websites to provide in-depth analysis of the technologies and vendors that websites adopt. Sincera focuses on adtech-related technologies and its clients include adtech companies. Sincera resembles Builtwith (`builtwith.com`), which identifies the technologies that websites use (Stroube & Dushnitsky, 2023) and has been used for economics research (e.g., Koning et al., 2022). However, Sincera's data is suited for deeper explorations of how websites use these technologies. For instance, Sincera illuminates websites' use of alternative user identifiers and their absorption by adtech intermediaries as well as websites' use of consent management platforms and compliance with California-compatible privacy consent signaling.[14]

Sincera regularly scans a list of websites to collect detailed data on its technology and vendor use. Sincera scanners use the Chrome browser to visit each website from within the United States, by default. Like Libert (2015), Sincera examines websites' use of HTML and JavaScript, though Sincera considers additional data like the content of a website's ads.txt file. As we elaborate in Section 3.2, Sincera scans hundreds of thousands of websites selected using a combination of client requests and filtering criteria.

Sincera's crawlers visit the site's homepage daily, but also scans additional pages on the domain. Sincera scans 5-10 additional pages of the site every 1-2 weeks. Sincera selects pages that are linked to on the homepage at random, though the crawler's logic seeks out new content and avoids index pages (e.g., `wired.com/authors`). This approach goes beyond most research that only scans the homepage (e.g., Libert, 2015). Karaj et al. (2018)

[14] See `https://sincera.io/blog/introducing-identifier-absorption` and `https://sincera.io/blog/who-s-afraid-of-ag-bonta`.

instead propose a methodology that relies on receiving data from a panel of users that also shows how websites operate beyond the homepage, though this sacrifices consistency due to the variability of user browsing over time. Sincera's baseline scans identify several site components including all interactions with first- and third-party domains and analyzing the page content (e.g., counts the number of images, text entries, videos, and ads on the page).

We engaged Sincera to collect data on how websites use Privacy Sandbox. Sincera is an ideal partner due to its expertise in extracting detailed data on websites' use of various technologies and vendors. In most cases, websites initiate Privacy Sandbox by calling APIs on the Chrome browser and passing standardized data types to the API. Sincera identifies sites that make these API calls and intercepts the data passed to the API. We ingest this raw data (i.e., often lengthy JSON) and extract relevant variables; notably, we identify the relevant adtech vendor from its domain name. The API data can be rich in principle, though the values often are encrypted so as not to reveal business information in practice. Sincera also collects other relevant data types: for instance, websites indicate their desire to opt out of Topics API in their HTTP header.

We focus on identifying which sites adopt which Privacy Sandbox technology as well as identifying the sites' adtech vendors where relevant. We currently track the Protected Audience API (i.e., buy- and sell sides) and Topics API (i.e., buy-side API calls and website opt-outs). This data collection begins between 2022 and 2023 as Sincera introduced the API detection capabilities sequentially. Our study focuses on whether we or not we detect a given API on a given website during a month, though our data collection could also allow to measure how often we detect the APIs.

Our data collection identifies websites baseline Privacy Sandbox adoption using a clean Chrome browser without any cookies. Our scanning approach has several limitations. First, we miss API calls by vendors that require that the user has a third-party cookie (e.g., storing consent). Second, we also miss or undercount vendors that exclusively or largely call the APIs for users who are labeled as part of the Google's Privacy Sandbox experiment. Chrome launched its voluntary Privacy Sandbox testing group for 7.5% of users in November 2023 and its Privacy Sandbox testing group for 0.75% of users on January 4, 2024.[15] These testing group signals were specifically developed for this experiment and are not straightforward to simulate for our data gathering purposes. So,

[15] In particular, the Mode A experiment allows vendors to voluntarily concentrate on Chrome users with the "label_only_*" experimental labels. The Mode B experiment creates two groups with third-party cookies deactivated: 0.75% of users with Privacy Sandbox APIs activated ("treatment_1.*" labels) and 0.25% of users without the APIs or cookies ("control_2" label). Chrome communicates these experimental group labels in two ways: via the HTTP header and via the `navigator.cookieDeprecationLabel.getValue()` JavaScript API. For more details, see CMA (2023b; 2023a) and `https://developers.google.com/privacy-sandbox/setup/web/chrome-facilitated-testing`.

we miss activity where a site or vendor only uses Privacy Sandbox APIs on users in the experimental groups. Third, we did not detect Protected Audience API use by Google's adtech until February 2024 for technical reasons. Specifically, Google's adtech checked the browser to determine the availability of Sandbox APIs before interacting with those APIs. Sincera adapted its scanners to respond favorably to these pre-screening requests in February 2024. Third, we do not observe API calls by vendors who block API calls originating from cloud data centers: i.e., rather than residential IP addresses. Fourth, we principally observe Privacy Sandbox adoption via JavaScript, so may miss vendors that use other means to access the APIs. In particular, we observe Topics API calls via JavaScript, but omit HTTP header calls to the API.

3.2 Website sample

Our website sample consists of a target list of 59,620 top sites. Sincera tracks a panel of over 250,000 sites and this panel has grown over time. Sincera chooses its site list using a combination of client requests and its own filtering criteria. Since the resulting list is idiosyncratic, we instead construct an explainable list of top sites that we draw from Sincera's panel.

Our sample begins with a Tranco list of top 100,000 websites drawn from the Alexa and Majestic site rankings. Le Pochat et al. (2019) propose the Tranco ranking to improve the reliability and reproducibility of website rankings for research purposes. Our subsequent filtering criteria is closely aligned with Sincera's filters. First, we exclude sites from certain countries outside the Americas and Europe using the site's top-level domain (e.g., ".ru", ".cn"). Second, we exclude certain non-commercial top-level domains like ".edu", ".gov", and ".org." Third, we exclude domains containing terms that are associated with low-quality sites like adult content, spam, and marketing. Fourth, we omit sites that are unreachable, return HTTP response errors, redirect to another domain, and certain domains belonging to web technology vendors.

Our final list consists of 59,620 top sites. As of April 2023, Sincera was scanning 52,546 of these websites in response to its customer demands. At our request and due to organic growth, Sincera's coverage of our target list grew over time: 56,885 of our sites (95.4%) by the end of 2023. We provide a detailed explanation of our filtering criteria and site sample construction in Appendix A.

4 Results: Privacy Sandbox

Our descriptive results present how websites and adtech companies use Privacy Sandbox technologies. Below, we consider the Protected Audience and Topics APIs.

4.1 Protected Audience API

As described in Section 2, the Protected Audiences API consists of buy-side and sell-side components. On the buy side, sites interact with the PA API to add the user to an interest group (or remove the user), so that an advertiser can show ads to that user on another website. On the sell side, sites that include advertising can interact with the PA API to run an auction to show a behaviorally targeted to the user based on that user's interest group membership. We consider both components of the PA API in turn below. As noted in Section 3.1, we do not observe Google's adtech use of PA API until February 2024.

4.1.1 PA interest groups

Adtech vendors play an important role in website adoption of PA interest groups. To facilitate testing, Chrome allows vendors that are present on a site's page to call the interest group PA API without requiring that site's permission. However, websites can opt out of PA API and Chrome will eventually require sites to explicitly allow PA API.[16] Though not required, adtech vendors may opt to notify websites and/or seek their permission before creating interest groups. In sum, we ascribe only a weak intent signal on behalf of websites when we detect an interest group PA API call on the site at present.

Figure 3 shows the evolution of the share of sites that add users to PA Interest Groups using the `joinAdInterestGroup` function. Between May and September 2023, this share held stable between 3.1% and 3.6% of sites. Adoption peaked in October 2024 at 4.7% of sites then fell to 1.2%-1.3% for the rest of 2023. As we explain below, this drop may reflect measurement issues rather than a true drop in PA adoption. The October increase could arise from the general availability of Privacy Sandbox APIs among Chrome users as well as increased vendor adoption.

By intercepting the data transmitted to the PA API, we can identify the adtech vendor that create the interest group via their associated web domain. Figure 4 shows the evolution of vendor market shares by site reach over time. For the first several months, Criteo had by far the largest market share followed by RTB House. Criteo appears to exit in November 2023, though we suggest below that this is a measurement issue. Adthrive

[16]Websites can disable PA site-wide by specifying a permission policy in their HTTP response header or specify this at the ad level within an iframe. As of 2023, we did not track PA opt-out choice. See `https://developers.google.com/privacy-sandbox/relevance/protected-audience-api/opt-out`.

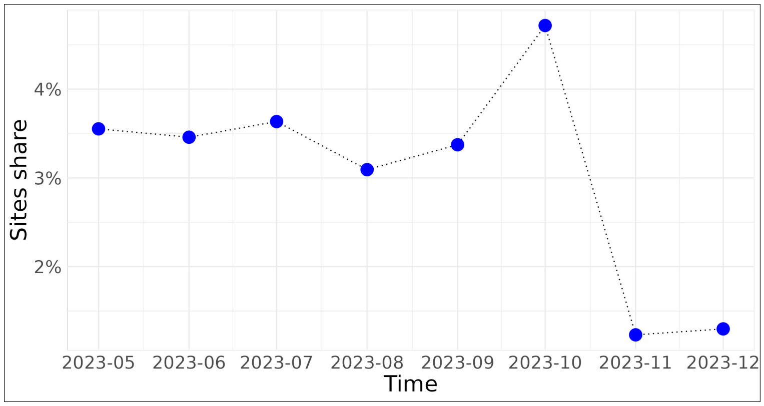


Figure 3: Protected Audience API Interest Groups: Site adoption over time

(now rebranded as Raptive) entered in September 2023 and became the dominant vendor by the end of 2023. Retargetly and Teads enter in October 2023, and though Retargetly appears to all but exit the next month. Some vendors experiment on a small scale: AdRoll, Amazon, and Seedtag all appear on a single site in our sample.

Figures 3 and 4 show reduced usage in November 2023, largely due to Criteo's apparent exit. We believe that these changes may arise from two technical issues, as discussed in Section 3.1. First, Criteo appears to filter out traffic originating from data centers rather than residential IP addresses. Sincera's crawlers originate from cloud data centers and appear to therefore omit Criteo. Second, November 2023 coincides with the beginning of Chrome's coordinated testing experiment. We only measure API use for unlabelled Chrome browsers. As such, the November drop may reflect vendors moving to the set of users in the voluntary API testing label group rather than a true reduction in PA API use. Nevertheless, the only vendor we observe exiting the 2023 data in November is Criteo, and we do not know whether the sandbox experiment contributed to this as well. We do however see notably reduced reach for both RTB House and Retargetly in November 2023 as well.

PA interest group API data (July 2023)

Below, we discuss the PA interest group data in greater depth using data from July 2023 and featuring wider sample of all sites tracked by Sincera at that time. We intercept PA interest group data, which provides some insight into ad buyer's preferences and how they use the PA API. Nevertheless, these signals are often masked or encrypted to protect the advertiser's and adtech vendor's business secrets. Notably, however, the "RenderURL" parameter can reveal the ad creative. We have not systematically investigated this data, but manual inspection suggests that these are predominantly retargeting ads where the creative often features multiple product images. Sincera detected 58,151 different PA

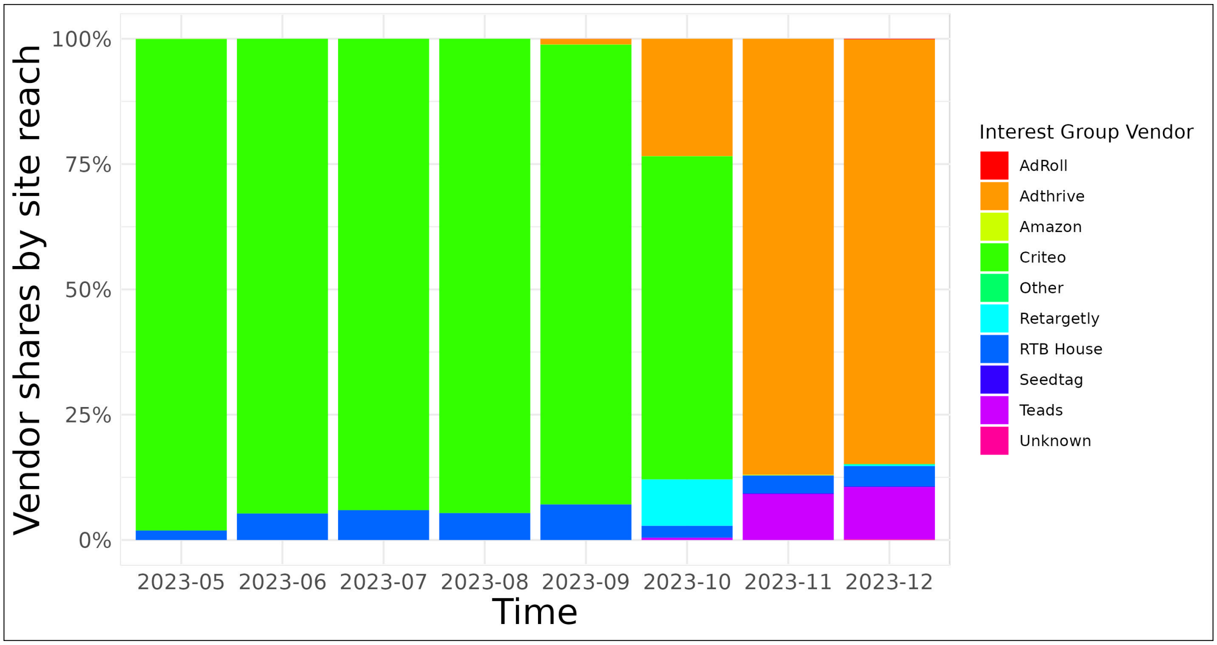


Figure 4: Protected Audience API Interest Groups: Adtech vendor share by distinct sites over time

Interest Groups being created on 4,042 publishers. This included 4,258 advertisers (according to the "name" field used by the API) yielding 4,521 unique publisher-advertiser combinations.

Only Criteo and RTB House are present at this time, and we use the detailed interest group data to compare how the two vendors use PA API. Note that Rumiński et al. (2022) detail RTB House's testing experience as of August 2022, when it also detected interest groups from Google (Doubleclick) and Criteo. On average, RTB House interest groups included 9.9 ad creative entries, but Criteo interest groups included only 0.8. Though vendors must specify an ad creative URL in this field, Criteo did not do so 49.5% of the time. Criteo explained that this was a deliberate choice that reflected testing behavior (Höring, 2022). Through the "adComponents" field, PA allows vendors to create ad creatives with subcomponents: e.g., dynamic retargeting creatives can use "adComponents" to feature different product images. Criteo did not use this functionality, but RTB House included an average of 194 such components (minimum of 9 and maximum of 1,400). RTB House always set the interest group duration of Chrome's maximum of 30 days, whereas Criteo always set the this to 43.2 minutes (perhaps erroneously).[17]

In terms of bidding, PA allows vendors to specify a numerical priority for the Interest Groups in cases where the seller restricts the number of interest groups that it considers

[17]The duration must be set in milliseconds, but Criteo has specified the duration to be the total seconds in 30 days.

per vendor in the PA auction. RTB House is specifying this parameter statically, while Criteo is not. Neither company used PA's optional parameters for dynamically calculating this priority parameter.[18] RTB House alone is using PA's option to run bidding functions in WebAssembly, rather than JavaScript. Finally, Criteo is passing a user ID as part of its userBiddingSignals parameters: presumably this is used for testing and debugging. In sum, Criteo is creating more PA interest groups, whereas RTB House appears to be using more of the technology's functionality.

4.1.2 PA API auctions

We also examine the sell side of the PA API, in which websites initiate a PA auction on the browser using the `runAdAuction` function.

From early data that we collected in June 2022, we detected PA API auctions on 1,363 distinct sites among the wider sample that Sincera scanned at the time (0.5% of the total). Unfortunately, we have not detected PA API auctions since then, despite scanning for this. As we noted in Section 3.1, this appears to arise from a technical issue whereby Sincera's scanners were not passing Google's browser pre-scanning for Sandbox APIs filters until February 2024. In November 2023, Google Ad Manager (GAM) also made clear that it would experiment with PA API auctions for all its publishers by default. Given that Google Ad Manager has a dominant position as a publisher ad server, most sites with ads would therefore be defaulted into PA API auctions. However, Google Ad Manager limits by default the share of PA API auctions it runs for Chrome users that are not assigned to the Google/CMA experiment.[19] Since our data collection employs an unlabeled Chrome user, we will observe fewer sites that employ GAM's PA API auctions then we would if our browser was part of the Privacy Sandbox experimental testing groups.

We therefore propose two conservative measurements of PA API adoption. First, we do observe PA API auctions for other sellers. In particular, we begin observing PA API auctions that specify Seedtag as the seller beginning in October 2023. Seedtag is the sole seller that we observe in our data as of the end of 2023. Moreover, Seedtag specifies itself as the unique PA buyer in all cases. Second, we also propose a another approach to detect PA auction adoption at the site level. Many websites sell advertising through a mechanism called header bidding using a free and open-source standard called Prebid. Prebid allows for websites to integrate PA auctions by adding a module called `fledgeForGpt` and enabling PA in their Prebid javascript configuration. Sincera scans websites and records their Prebid configuration, which allows us to identify sites that have enabled

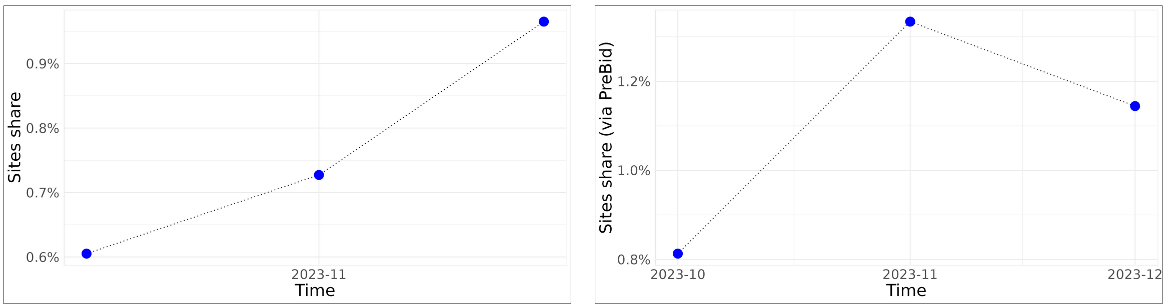
[18] These are the priorityVector, prioritySignalsOverrides, and enableBiddingSignalsPrioritization parameters.

[19] `https://support.google.com/admanager/answer/13178817`.

Figure 5: Protected Audience API auctions: Site adoption over time

(a) Detected by API interactions (non-Google sellers)

(b) Detected using site's Prebid settings



fledgeForGpt. However, since this measurement is specific to Prebid, we omit sites that adopt PA API auctions but do not use Prebid.

Figure 5 plots both (conservative) measures of PA auction adoption. Figure 5a plots the share of websites engaging non-Google PA sellers. In 2023, these are exclusively auctions with Seedtag as the seller (and exclusively buyer). This grew from 0.61% of tracked sites in October 2023 to 0.97% at the end of 2023. Figure 5b plots the share of websites that integrated PA auctions into their Prebid configuration. Website adoption increased from 0.8% in October to 1.3% in November before falling to 1.1% in December 2023. In December 2023, the sets of users that use Seedtag as a PA seller or specify PA auctions in Prebid almost does not overlap—representing 2.1% of tracked websites.

PA API auction data (July 2022)

We intercept the PA auction API data in June 2022 across all sites tracked by Sincera at the time. As with the buy side, the raw JSON data provides some insight into this ad market while masking and encrypting several business-sensitive inputs. We note that the data identify partners that can be included in the PA auction. However, adtech vendors need not be identified in the PA auction as buyers to participate in the auction, so this undercounts vendor participation. Nevertheless, the PA auction API data most often lists Google alone (67.7% of cases) or RTB House alone (30.5%) as buyers,[20] but lists both in 1.0% of cases and no sellers in 0.7% of cases. Despite this, Google Ad Manager sends auction signals to its buy side (see below) in all cases.

The auction data also identify specific data that are passed to individual vendor participants under the `perBuyerSignals` input to the PA API. We find that these Google's sell-

[20] RTB House appears as three domains, reflecting servers in different locations throughout the world.

side initiated auctions pass less volume of data on average to RTB House (3.7 kilobytes) than to Google's buy side (10.0 kilobytes). However, these may not be directly comparable due to different encoding and different data structures. RTB House uses a JSON structure with 18 data elements, some of which may be interpretable (e.g., time stamps, ad size format, URLs). Google instead uses a nested array consisting of numbers.[21]

4.2 Topics API

We first consider the Topics API labels that Google assigns to each the domains in our target set. To generate this correspondence, we use Chrome's on-browser classifier (`chrome://topics-internals`) and use version 2 of the Topics taxonomy. Figure 7a shows a histogram of the number of labels per domain. The median site domain has a single label while 8.35% of sites are not classified. The mean site has 1.37 labels and the site with the most labels (8) is `huawei.com`. Figure 7b presents the most frequently appearing labels, however, this uses version 1 of the taxonomy and all sites tracked by Sincera.

As described in Section 2, the Topics API consists of vendor calls on sites to request user segment information. Moreover, publishers can choose to opt out of Topics (i.e., both providing and receiving data). We consider both actions in turn below.

4.2.1 Topics API Calls

We record JavaScript calls to the Topics API requesting the behavioral segments associated with a given user. As noted in Section 3.1, we do not observe the alternative mechanisms for calling the Topics API: i.e., via HTTP Header. The HTTP header approach can be accessed via fetch request or via an iframe attribute.[22] Chrome's platforms statistics indicate that it detects Topics API calls via HTTP fetch requests on 6.95% of page loads and via iframe attributes on 2.82% of page loads as of March 2024.[23]

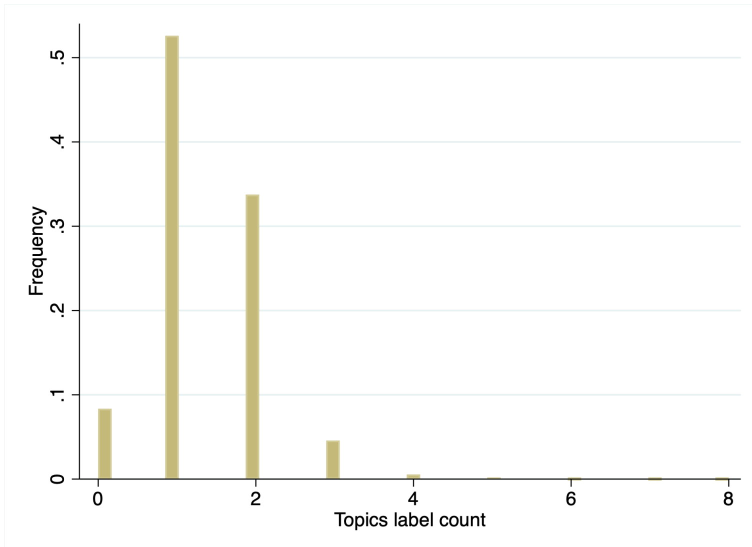
The Topics API interactions are simpler than the PA API in many ways. Adtech vendors present on the page can call the Topics API without requiring additional permission from the website. The API returns up to three topics to the vendor, which the vendor can use for ad bidding. This simplicity favors broader adoption and Figure 7 shows that the share of sites with Topics API calls has grown over time from 25.1% of sites in October to 36.2% of sites in December 2023.

[21]Note that this data structure includes a large number of space characters for formatting, which appears to artificially increase the data size (e.g., by a factor of 4 in one case).

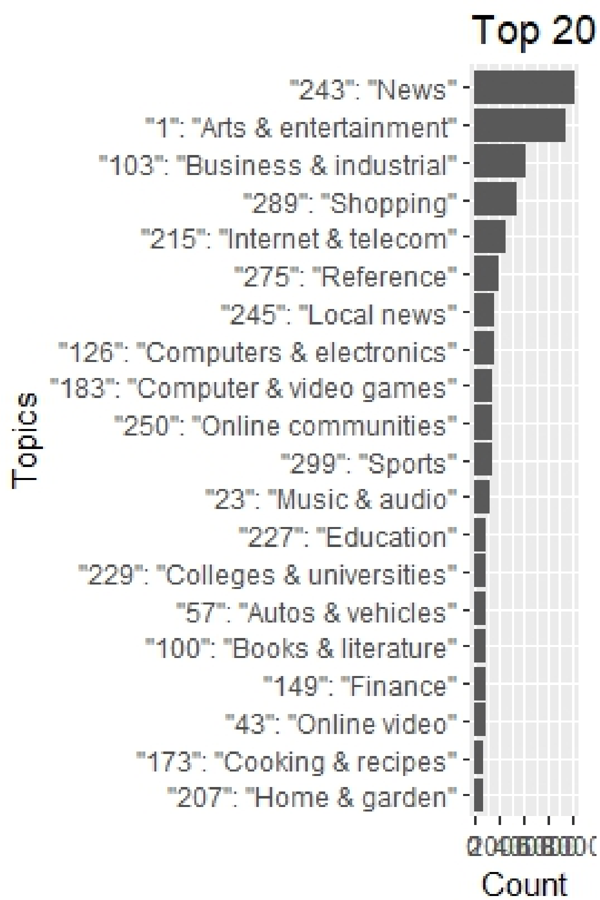
[22]See Google's developer guide for details: `https://developers.google.com/privacy-sandbox/relevance/topics/developer-guide`

[23]Monthly statistics for each are provided at `https://chromestatus.com/metrics/feature/timeline/popularity/4460` and `https://chromestatus.com/metrics/feature/timeline/popularity/4497`, respectively.

Figure 6: Topics API labels assigned to target domain list

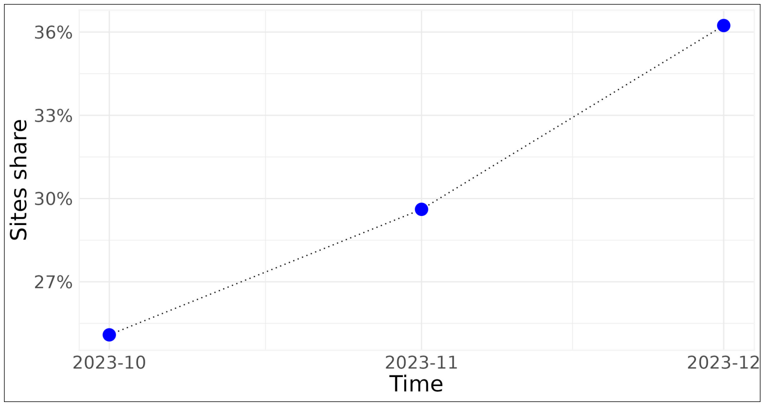


(a) Number of labels per domain



(b)

Figure 7: Topics calls (via JavaScript): Site adoption over time



Our Topics API figures are conservative because we only measure JavaScript calls to API. In particular, Google added a second way for vendors to call the Topics API, i.e., via HTTP header that we do not observe.[24] Note that we lack information on the segment data returned by the Topics API (see instead Król, 2023; Selman, 2022), which can only be viewed by third-party domains that are called on multiple domains.

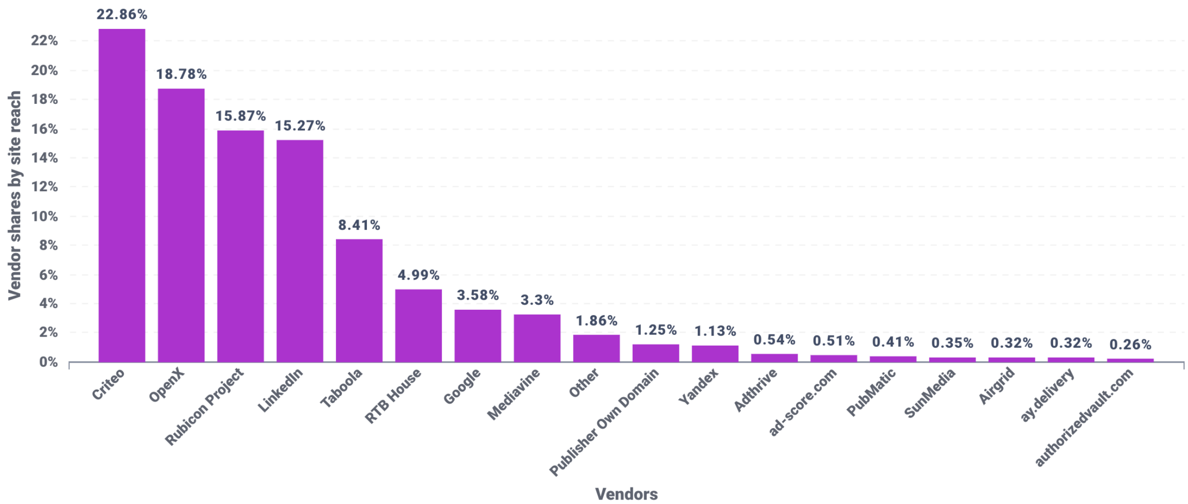
Using the data in Topics API call, we observe the adtech vendor's domains from which we can infer the vendor's identity. Figure 8 shows adtech vendor market shares by website for our most recent month observation (December 2023). We observe more adtech vendors in this data than with PA Interest Groups (Figure 4) and the largest vendors include: Criteo, OpenX, Rubicon Project, LinkedIn, Taboola, RTB House, Google, and Mediavine. We also see websites (e.g., `dailymail.co.uk`, `usatoday.com`) experimenting with Topics calls on their own site or network of sites. We group these under the "Publisher Own Domain" heading, which collectively represent 1.25% of website-vendor pairs. However, since Topics only returns data for domains observed on that category's page, such experiments may return little user segment data if any. We also group vendors that appear on fewer than 100 sites as "Other", which collectively represent 1.9% market share.

Note that the Topics API call source consists of a URL that often passes additional parameters to that domain such as the publisher ID and consent-related data.[25] However, these parameters are not required by the Topics API, so vendors choose which data to include and how to format this for their own purposes. In July 2023 and across all sites scanned by Sincera, Criteo in particular includes a lot of data in its call source APIs, often

[24] For related documentation, see `https://developers.google.com/privacy-sandbox/relevance/topics/developer-guide`. Based on discussions with Google's Topics API team, the HTTP header method represents a growing share of Topics API calls.

[25] E.g., `https://gum.criteo.com/syncframe?origin=publishertag&topUrl=www.domain.com&us_privacy=1YN-`

Figure 8: Topics calls: Vendor site reach (December 2023)



including lengthy consent-related strings: a median of 542 characters and a mean of 633.3 characters in the URL after its domain name. In comparison, the other vendors include a median of 36 characters and mean of 42.0 characters after their domain name in the call source URLs.

4.2.2 Website opt-outs

Websites can opt out of participating with Topics API by including an opt-out flag in their HTTP header. Specifically, websites declare this in their "Permissions Policy."[26] In this case, the Topics API omits this domain when the API assigns users to behavioral segments. Sites that opt out of Topics API, however, can not receive data from the Topics API either. So, these sites forgo any benefit or cost from participating in the Topics API.

Figure 9 plots the opt-out share over time. This fell from 1.6% in September to 1.0% in December 2023. We have earlier data from July 2022, which allows us to bound the opt-out rate between 2.7% and 3.9% at that time.[27] This decline may arise from Google shelving its more controversial Federated Learning of Cohorts (FLoC) proposal in 2022 and replacing it with Topics API. FLoC resembles Topics API in that it assigns users to behavioral segments. However, FLoC allowed for thousands of (unlabeled) user behavioral

[26]Permissions Policy "provides mechanisms for web developers to explicitly declare what functionality can and cannot be used on a website" (`https://developer.mozilla.org/en-US/docs/Web/HTTP/Permissions_Policy`).

[27]Sincera's historical data is limiting because it is not structured as panel data. Rather, Sincera only records when it first records a site engaging in a tracked behavior and how often it has recorded that behavior since. We therefore bound the opt-out rate in July 2022 from below by the total number of sites that Sincera only ever recorded as opting out and from above by all those sites that Sincera observed ever opting out until that date.

Figure 9: Topics Opt Out: Evolution over time

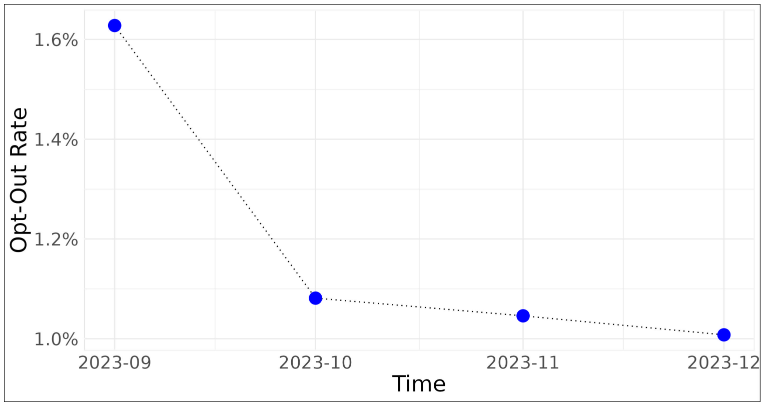
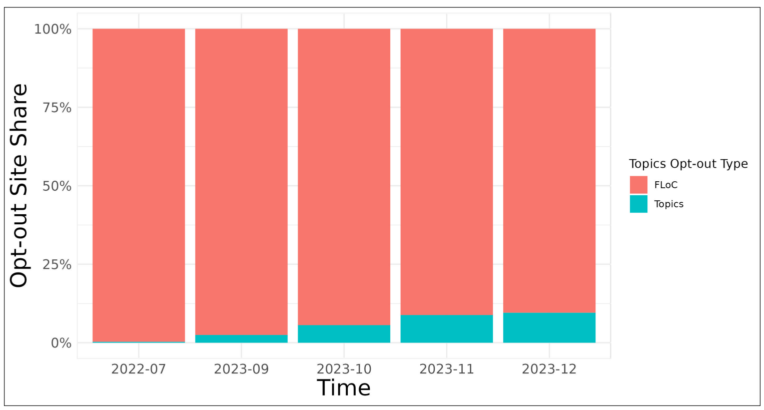


Figure 10: Topics Opt Out: Evolution of opt-out type share



segments, which drew criticism for its potential privacy risk.

Google uses the same opt-out flag for both Topics and FLoC, though websites could specify one or the other. Sites did not need to update this flag, however, as Google announced that Topics API would respect both FLoC and Topics opt-out signals. Figure 10 shows the evolution of opt-out type by share of websites. In July 2022, almost all of these opt-out flags (over 99%) refer to FLoC. By the end of 2023, the share of sites specifying an opt-out signal specific to Topics API grew to 9.6% of opt-out sites. Combined, this suggests that websites were making active choices about participating in Topics/FLoC and updating their choices in response to the Google's evolving proposals.

5 Conclusion

We contribute a first study of the online ad industry's adoption of privacy-enhancing technologies. We focus on Protected Audience and Topics APIs proposed by Google as part of its Privacy Sandbox initiative. We devise a method for detecting these technologies on a website and intercepting their associated data. Topics API stands out for its greater adoption both by sites (36.2% in December 2023) and adtech vendors—perhaps due to its greater ease of use. Websites' opt out rate for Topics API has decreased over time to only 1.0% at the end of 2023. On the advertiser side, adoption of Protected Audience API peaked at 4.7% of sites in October. However, new vendors entered in the last months of 2023, while the largest vendor (Criteo) appeared to exit—perhaps focusing on users in Chrome's Privacy Sandbox experimental groups. On the publisher side, about 1% of sites allow Protected Audience API ad auctions as indicated by their header bidding (i.e., Prebid) settings.

Our research has several limitations. Our measurement methodology is traditional in the sense that we use clean Chrome browsers without cookies or experimental labels. Consequently, we miss vendors that require cookies to query the Sandbox APIs (e.g., Google's adtech) as well as vendors that only buy Chrome users with Privacy Sandbox experimental labels. For technical reasons, we do observe Google's adtech use of PA API until February 2024 and we only observe Criteo's use until October 2023. We observe API use via Javascript but do not observe alternatives: we omit HTTP header calls to the Topics API in particular. Ours is a measurement study, so we lack data on users or website outcomes. We leave the consequences of Privacy Sandbox for users and websites to future research.

In the future, we will continue to collect data. We maintain a public dashboard (`app.sincera.io/privacysandbox`) that provides adoption statistics. We anticipate that the milestones on Google Chrome's roadmap to cookie deprecation will propel greater adoption of Privacy Sandbox APIs. We also intend to investigate which websites choose to adopt or opt-out of Privacy Sandbox APIs.

References

Acquisti, A. (2024). The economics of privacy at a crossroads. In A. Goldfarb & C. Tucker (Eds.), *The Economics of Privacy* chapter 2. University of Chicago Press.

Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature*, 54(2), 442–92.

Ada, S., Abou Nabout, N., & Feit, E. M. (2022). Context information can increase revenue in online display advertising auctions: Evidence from a policy change. *Journal of Marketing Research*, 59(5), 1040–1058.

Akbarpour, M. & Li, S. (2020). Credible auctions: A trilemma. *Econometrica*, 88(2), 425–467.

Alcobendas, M., Kobayashi, S., Shi, K., & Shum, M. (2023). The impact of privacy measures on online advertising markets. Available at SSRN 3782889.

Alvim, M. S., Fernandes, N., McIver, A., & Nunes, G. H. (2023). A quantitative information flow analysis of the topics api. In *Proceedings of the 22nd Workshop on Privacy in the Electronic Society*, WPES '23 (pp. 123–127). New York, NY, USA: Association for Computing Machinery.

Anand, P. & Lee, C. (2023). Using deep learning to overcome privacy and scalability issues in customer data transfer. *Marketing Science*, 42(1), 189–207.

Aridor, G. & Che, Y.-K. (2024). Privacy regulation and targeted advertising: Evidence from apple's app tracking transparency. *Available at SSRN*.

Bleier, A., Goldfarb, A., & Tuckerc, C. (2020). Consumer privacy and the future of data-based innovation and marketing. *International Journal of Research in Marketing*.

Carey, C., Dick, T., Epasto, A., Javanmard, A., Karlin, J., Kumar, S., Medina, A. M., Mirrokni, V., Nunes, G. H., Vassilvitskii, S., & Zhong, P. (2023). Measuring re-identification risk.

Cecere, G. & Lemaire, S. (2023). Have i seen you before? measuring the value of tracking for digital advertising. Available at SSRN.

Choi, H., Mela, C., Balseiro, S., & Leary, A. (2020). Online display advertising markets: A literature review and future directions. *Information Systems Research*, 31(2), 556–575.

Choi, W. J. & Jerath, K. (2022). Privacy and consumer empowerment in online advertising. *Foundations and Trends® in Marketing*, 15(3), 153–212.

Competition and Markets Authority (2023a). *Quantitative testing of Google's Privacy Sandbox technologies – Additional CMA guidance to third parties on testing*. Technical report, Competition and Markets Authority.

Competition and Markets Authority (2023b). *Quantitative testing of Google's Privacy Sandbox technologies – CMA guidance to third parties on testing*. Technical report, Competition and Markets Authority.

Cooper, D. A., Yalcin, T., Nistor, C., Macrini, M., & Pehlivan, E. (2023). Privacy considerations for online advertising: A stakeholder's perspective to programmatic advertising. *Journal of Consumer Marketing*, 40(2), 235–247.

David, P. A. & Greenstein, S. (1990). The economics of compatibility standards: An introduction to recent research 1. *Economics of Innovation and New Technology*, 1(1-2), 3–41.

Englehardt, S. & Narayanan, A. (2016). Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1388–1401).: ACM.

Epasto, A., Medina, A. M., Ilvento, C., & Karlin, J. (2022). Measures of cross-site reidentification risk: An analysis of the topics api proposal.

Farahat, A. & Bailey, M. C. (2012). How effective is targeted advertising? In *Proceedings of the 21st international conference on World Wide Web* (pp. 111–120).: ACM.

Forman, C., Goldfarb, A., & Greenstein, S. (2005). Geographic location and the diffusion of internet technology. *Electronic Commerce Research and Applications*, 4(1), 1–13.

Geradin, D., Katsifis, D., & Karanikioti, T. (2021a). Google as a de facto privacy regulator: analysing the privacy sandbox from an antitrust perspective. *European Competition Journal*, 17(3), 617–681.

Geradin, D., Katsifis, D., & Karanikioti, T. (2021b). Google as a de facto privacy regulator: analysing the privacy sandbox from an antitrust perspective. *European Competition Journal*, 17(3), 617–681.

Goldfarb, A. (2013). What is different about online advertising? *Review of Industrial Organization*, (pp. 1–15).

Goldfarb, A. & Que, V. F. (2023). *The Economics of Digital Privacy*. Working Paper 30943, National Bureau of Economic Research.

Goldfarb, A. & Tucker, C. (2011). Privacy regulation and online advertising. *Management Science*, 57(1), 57–71.

Goldfarb, A. & Tucker, C. E. (2024). *The Economics of Privacy*. University of Chicago Press.

Google (2022). The Privacy Sandbox: technology for a more private web. https://privacysandbox.com.

Google (2023). *Google Ads Test: Testing Interest Based Audience Solutions with Privacy Preserving Signals*. Technical report, Google.

Gordon, B. R., Jerath, K., Katona, Z., Narayanan, S., Shin, J., & Wilbur, K. C. (2021). Inefficiencies in digital advertising markets. *Journal of Marketing*, 85(1), 7–25.

Höring, F. (2022). An update on fledge chrome testing. Criteo R&D Blog.

Information Commissioner's Office (2022). Privacy-enhancing technologies (pets). In *Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance* chapter 5. Information Commissioner's Office.

Janssen, R., Kesler, R., Kummer, M. E., & Waldfogel, J. (2022). GDPR and the lost generation of innovative apps. NBER working paper.

Jerath, K. & Miller, K. M. (2023). Using the dual-privacy framework to understand consumers' perceived privacy violations under different firm practices in online advertising. Available at SSRN.

Johnson, G. A., Lin, T., Cooper, J., & Zhong, L. (2023a). COPPAcalypse? The YouTube settlement's impact on kids content. Available at SSRN 4430334.

Johnson, G. A., Runge, J., & Seufert, E. (2022). Privacy-centric digital advertising: Implications for research. *Customer Needs and Solutions*, 9(1), 49–54.

Johnson, G. A., Shriver, S., & Goldberg, S. (2023b). Privacy & market concentration: Intended & unintended consequences of the GDPR. *Management Science*, 69(10), 5695–6415.

Johnson, G. A., Shriver, S., & Goldberg, S. (2023c). Privacy & market concentration: Intended & unintended consequences of the GDPR. *Management Science*, (pp. to appear).

Johnson, G. A., Shriver, S. K., & Du, S. (2020). Consumer privacy choice in online advertising: Who opts out and at what cost to industry? *Marketing Science*, 39(1), 33–51.

Karaj, A., Macbeth, S., Berson, R., & Pujol, J. M. (2018). Whotracks.me: Monitoring the online tracking landscape at scale. *CoRR*, abs/1804.08959.

Kircher, T. & Foerderer, J. (2024). Ban targeted advertising? an empirical investigation of the consequences for app development. *Management Science*, 70(2), 1070–1092.

Koning, R., Hasan, S., & Chatterji, A. (2022). Experimentation and start-up performance: Evidence from A/B testing. *Management Science*, 68(9), 6434–6453.

Kraft, L., Skiera, B., & Koschella, T. (2023). Economic impact of opt-in versus opt-out requirements for personal data usage: The case of apple's app tracking transparency (att). Available at SSRN 4598472.

Król, M. (2023). Topics api today (december 2023): Assessing its scale and usability. RTB House Blog.

Lachner, F., Chen, Y., & Olsauskas-Warren, T. (2023). User attitudes towards controls for ad interests estimated on-device by the browser. In *Symposium on Usable Security and Privacy (USEC)*.

Le Pochat, V., Van Goethem, T., Tajalizadehkhoob, S., Korczyński, M., & Joosen, W. (2019). Tranco: A research-oriented top sites ranking hardened against manipulation. In *Proceedings of the 26th Annual Network and Distributed System Security Symposium*, NDSS 2019.

Lefrere, V., Warberg, L., Cheyre, C., Marotta, V., & Acquisti, A. (2022). Does privacy regulation harm content providers? A longitudinal analysis of the impact of the GDPR. Available at SSRN.

Lerner, A., Simpson, A. K., Kohno, T., & Roesner, F. (2016). Internet Jones and the raiders of the lost trackers: An archaeological study of web tracking from 1996 to 2016. In *25th USENIX Security Symposium (USENIX Security 16)*.

Li, D. & Tsai, H.-T. (2023). Mobile apps and targeted advertising: Competitive effects of data sharing. *Available at SSRN 4088166*.

Li, S., Schneider, M. J., Yu, Y., & Gupta, S. (2022). Reidentification risk in panel data: Protecting for k-anonymity. *Information Systems Research*, 34(3), 1066–1088.

Libert, T. (2015). Exposing the hidden web: An analysis of third-party HTTP requests on one million websites. *International Journal of Communication*, 9, 18.

Lukic, K., Miller, K. M., & Skiera, B. (2023). The impact of the General Data Protection Regulation (GDPR) on the amount of online tracking. Available at SSRN 4399388.

Martin, K., Nissenbaum, H., & Shmatikov, V. (2023). No cookies for you!: Evaluating the promises of big tech's 'privacy-enhancing'techniques. *Evaluating The Promises Of Big Tech's 'Privacy-Enhancing'Techniques.(December 9, 2023)*.

McGuigan, L., Sivan-Sevilla, I., Parham, P., & Shvartzshnaider, Y. (2024). Private attributes: The meanings and mechanisms of "privacy-preserving" adtech. *New Media & Society*, forthcoming, 14614448231213267.

Miller, K. M. & Skiera, B. (2023). Economic consequences of online tracking restrictions: Evidence from cookies. *International Journal of Research in Marketing*.

Neumann, N., Tucker, C. E., Subramanyam, K., & Marshall, J. (2023). Is first-or third-party audience data more effective for reaching the 'right'customers? the case of it decision-makers. *Quantitative Marketing and Economics*, 21(4), 519–571.

Neumann, N., Tucker, C. E., & Whitfield, T. (2019). Frontiers: How effective is third-party consumer profiling and audience delivery?: Evidence from field studies. *Marketing Science*, 38(6), 918–926.

Nikiforakis, N., Kapravelos, A., Joosen, W., Kruegel, C., Piessens, F., & Vigna, G. (2013). Cookieless monster: Exploring the ecosystem of web-based device fingerprinting. In *Security and privacy (SP), 2013 IEEE symposium on* (pp. 541–555).: IEEE.

Nottingham, M. (2021). Playing fair in the privacy sandbox: Competition, privacy and interoperability standards. Available at SSRN.

Olejnik, L. (2023). *Handbook on the Politics and Governance of Big Data and Artificial Intelligence*, chapter On the governance of privacy-preserving systems for the web: should Privacy Sandbox be governed? Elgar Handbooks in Political Science. Edward Elgar Publishing.

Olejnik, L. (2024). *Reconciling Privacy Sandbox initiatives with EU data protection laws*. PhD thesis, University of Edinburgh Law School.

Peukert, C., Bechtold, S., Batikas, M., & Kretschmer, T. (2022). Regulatory spillovers and data governance: Evidence from the GDPR. *Marketing Science*, 41(4), 318–340.

Quéré, R. (2022). On the topic of Topics. Xandr Tech Blog.

Rafieian, O. & Yoganarasimhan, H. (2021). Targeting and privacy in mobile advertising. *Marketing Science*, 40(2), 193–218.

Ravichandran, D. & Korula, N. (2019). *Effect of disabling third-party cookies on publisher revenue*. Technical report, Google Inc.

Rumiński, M., Iwańczak, P., & Włodarczyk, Ł. (2022). Findings from the early fledge experiments. *Available at SSRN 4219796*.

Schneider, M. J., Jagpal, S., Gupta, S., Li, S., & Yu, Y. (2018). A flexible method for protecting marketing data: An application to point-of-sale data. *Marketing Science*, 37(1), 153–171.

Selman, E. (2022). Topics api: Criteo's first look at google's interest-based advertising solution. Criteo R&D Blog.

Sokol, D. D. & Zhu, F. (2021). Harming competition and consumers under the guise of protecting privacy: An analysis of Apple's iOS 14 policy updates. *Available at SSRN*.

Stroube, B. & Dushnitsky, G. (2023). Built with BuiltWith: A technical note. *Available at SSRN*.

Technology Analysis Division of the Office of the Privacy Commissioner of Canada (2017). *Privacy Enhancing Technologies – A Review of Tools and Techniques*. Technical report, Office of the Privacy Commissioner of Canada.

Thomson, M. (2023). A privacy analysis of google's topics proposal.

Tucker, C. (2024). The economics of privacy: An agenda. In A. Goldfarb & C. Tucker (Eds.), *The Economics of Privacy* chapter 1. University of Chicago Press.

Tucker, C. E. (2012). The economics of advertising and privacy. *International Journal of Industrial Organization*, 30(3), 326 – 329. Selected Papers, European Association for Research in Industrial Economics 38th Annual Conference, Stockholm, Sweden, September 1-3, 2011.

Wernerfelt, N., Tuchman, A., Shapiro, B., & Moakler, R. (2022). Estimating the value of offsite data to advertisers on meta. Available at SSRN: https://ssrn.com/abstract=4176208.

Zolas, N., Kroff, Z., Brynjolfsson, E., McElheran, K., Beede, D. N., Buffington, C., Goldschlag, N., Foster, L., & Dinlersoz, E. (2021). Advanced technologies adoption and use by US firms: Evidence from the annual business survey. NBER working paper.

Appendix

A Website sample construction

We want to track a consistent and explainable list of higher quality, commercial websites over time. Constructing our target website list was an iterative process as some of Sincera's Privacy Sandbox data collection predates these efforts. Once we finalized our target list, Sincera agreed to add the remaining sites to its roster of regularly scanned sites, so our sample of scanned sites grows over time.

We begin with a Tranco list of top 100,000 sites as of October 2022, which is designed and constructed using the methodology from Le Pochat et al. (2019). Our chosen Tranco list aggregates independent rankings by Alexa and Majestic: the list is available at `https://tranco-list.eu/list/4K8VX/1000000`.[28] We chose the 100,000 top sites threshold in order balance our site coverage and the overlap with Sincera's regularly scanned sites as of mid-2023. From this, we constructed a list of 59,620 target sites after filtering out certain domain names and removing unscannable websites. Sincera's initial coverage of our target site list was high (88.8%) in May 2023 and grew to 95.4% by the end of 2023.

Section A.1 describes our *ex ante* website domain filtering criteria to focus on higher quality, commercial domains. Section A.2 describes our *ex post* list of scannable websites that removes unreachable, redirecting, or web-vendor domains. Section A.2 concludes by describing how our data collector's scanning coverage of our target list grows over time.

A.1 Domain filtering criteria

Sincera is a startup that regularly scans a list of websites to collect data on the site's content as well as the vendors and technologies that the site employs. Sincera scans hundreds of thousands of websites every day and this list has grown over time. Sincera considers several factors in selecting the sites on its scan list, however the most important criterion is customer demand. Sincera is a commercial service whose customers can request scans of sites for a list that the customer provides. Sincera's customers include adtech companies, web publishers, and even journalists. As a startup, Sincera applied an initial criteria to filter domains opting to prioritize commercially-relevant websites and cost efficiency over full coverage. Nevertheless, customer demands override Sincera's domain filtering

[28] The primary Tranco list aggregates independent rankings from Umbrella and Farsight in addition to Alexa and Majestic (Le Pochat et al., 2019). In practice, this list places a greater emphasis on network traffic than consumer traffic alone. For instance, its top 10 list include domains that would be unfamiliar to many consumers like `gtld-servers.net` (#2), `akamaiedge.net` (#5), `amazonaws.com` (#10). We therefore prefer to use our (Alexa & Majestic) Tranco list, which also includes more sites that Sincera tracked in mid-2023.

criteria.

Before scanning websites, we filter our Tranco list of domains using three broad criteria that closely align with Sincera's initial filtering criteria. First, we exclude domains whose top-level domain indicates that the site is located in certain countries (e.g., `.ru`, `.cn`). Sincera's excluded country list is *ad hoc*, but predominantly includes large countries in Asia, Africa, and the Middle East. The full country list is: Angola, Azerbaijan, China, India, Iran, Japan, Kazakhstan, Kenya, Laos, Morocco, North Macedonia, Russia, Saudi Arabia, Singapore, South Korea, Taiwan, Tanzania, Turkey, Ukraine, Uzbekistan, and Vietnam. Relatedly, we exclude country-specific instances of `google.com` (e.g., `google.ro`) since these often redirect to `google.com` when scanned from outside that country.

Second, Sincera excludes certain non-commercial, top-level domains including: `.edu`, `.gov`, `.land` (i.e., real estate), `.mil` (i.e., military), `.org`, and `.win` (i.e., Microsoft Windows).

Third, Sincera seeks to exclude low-quality sites by excluding domains containing certain terms (including the top-level domain). These terms fall into the following broad categories:

- Adult content.[29]
- Spam content including the terms: casino, cash, free, game, mp3, pharma, torrent, and yeezy.
- Marketing content including the terms: ads, advertising, agency, cdn, digital, dns, group, marketing, media, pub, publisher, seo, solutions, and yield.
- Domains containing "blogspot": this removes sites from Google's popular blogging service `blogger.com`. Advertising on these sites is optional, and is restricted to Google AdSense ads.
- Domains containing "desi": this predominantly avoids content associated with the countries India, Pakistan, or Bangladesh including websites with the top-level domain "`.desi`".

We acknowledge that some of these terms are broad; for instance, the "desi" criterion also excludes domains containing the term "design." Nonetheless, our filtered domain list balances Sincera's scanning coverage with our goals of an explainable list of quality, commercially-relevant websites. These domain filtering criteria eliminate 24.2% of sites in our Tranco list, leaving an interim list of 75,797 sites.

[29]The excluded adult-rated content terms include: adult, sugardaddy, viagra, brides, nsfw, fetish, tranny, slut, porn, fuck, sex, and xxx.

A.2 Scannable websites

From the interim list of 75,797 sites, we find that 59,640 of these sites are scannable. In particular, we remove sites that are unreachable, generate HTTP errors, are web vendor domains, or redirect to another domain.

We built our own site-scanning algorithm to evaluate domains on our list that Sincera did not scan. Surprisingly, our Tranco list includes many domains that are unreachable or redirect to other domains. Our site scanning algorithm evaluates whether a site is reachable, records the HTTP response code returned by the domain, and records any redirect domain.

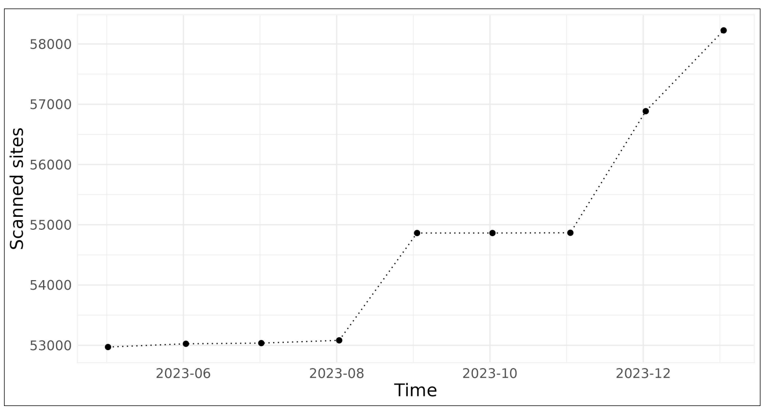
We classify unscannable websites using five criteria, which we apply cumulatively. First, our site scanner identifies 11,423 websites that are unreachable. We expect that many of these sites are not consumer-facing sites: e.g., web technology vendors employ these domains to provide advertising and content services for websites. Second, another 482 websites are unreachable in that they return HTTP codes that indicate either a client- or server-side error. Third, we eliminate 3,488 sites that redirect to different domains; however, many of these redirect domains are included in our list (e.g., `a.co` redirects to `amazon.com`). Fourth, we eliminate 403 domains that are identified by Sincera as associated with web technology vendors, and that Sincera does not scan. Fifth, we requested that Sincera add any remaining sites from our target list to Sincera's regular scanning list in August 2023. At this time, Sincera identified another 361 sites that it could not scan for various reasons.[30]

After these steps, we have a target list of 59,640 filtered, scannable websites. Sincera's site scanning coverage for our target list has grown over time. Sincera's initial coverage of our filtered list was high: 52,971 sites (88.8%) in May 2023. The missing sites are disproportionately lower ranked.[31] At our request, Sincera began adding the missing sites from our target list to its regular scanning list. Figure A.1 plots the growth of the number of scanned sites from our target list. By the end of 2023, Sincera was scanning 56,885 sites on our list (95.4%) and 58,226 sites by the middle of January 2024.

[30] Sincera encountered HTTP errors on 155 sites, Captchas on 130 (i.e., blocking automated scanning), failed page loads for 69, general errors for 6, and the scan timed out for 1 site.

[31] Sincera uses the primary Tranco site list (see footnote 28) to guide its scanning decisions. Half of the unscanned sites as of December 2023 are ranked outside of the primary Tranco top 250,000 list, which contributed to Sincera ignoring these sites.

Figure A.1: Number of Sincera scanned sites from our target site list



Note: This plots the number of sites that Sincera scanned as of the end of a given month from our target list of 59,640 filtered, scannable websites.